

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: November 29, 2004, 08:19:58 ; Search time 31.6873 Seconds  
(without alignments)  
2596.165 Million cell updates/sec

Title: US-09-936-333-27

Perfect score: 4683  
Sequence: 1 MGSDRRARKGGGGRDAGL.....PGVYTRLPFRDMKENTGV 855

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database : PIR-79.\*

1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3866	82.6	855	2 JC7731	membrane-bound arg
2	724	15.5	1034	1 A53663	enteropeptidase (E
3	709.5	15.2	1035	1 A43090	enteropeptidase (E
4	679.5	14.5	1019	1 A56318	enteropeptidase (E
5	661.5	14.1	1113	2 JE0315	low-density lipopr
6	576.5	12.3	1524	2 T30337	polyprotein - Altr
7	570	12.2	699	1 I54763	Ra-reactive factor
8	528.5	11.3	638	1 KQMSPL	plasma kallikrein
9	507.5	10.8	638	1 KORTPL	plasma kallikrein
10	506.5	10.8	790	1 PLRG	plasma (EC 3.4.21
11	504	10.8	613	2 S15468	complement C3b/C4b
12	495.5	10.6	786	1 A17547	serine proteinase
13	495	10.6	638	1 KQHUP	plasma kallikrein
14	493	10.4	583	2 B61545	plasma (EC 3.4.21
15	488.5	10.4	583	2 A29154	complement factor
16	488	10.4	810	1 PLHU	plasma (EC 3.4.21
17	486	10.4	416	1 KRBQ	coagulation factor
18	486	10.4	417	1 S00845	hepsin (EC 3.4.21.
19	485.5	10.4	812	1 PLBO	plasma (EC 3.4.21
20	483.5	10.3	343	1 A57014	proctasin (EC 3.4.
21	483.5	10.3	812	1 PLMS	plasma (EC 3.4.21
22	479.5	10.2	416	1 S33777	hepsin (EC 3.4.21.
23	477	10.2	810	2 B30848	plasma (EC 3.4.21
24	475	10.1	686	1 A59271	Ra-reactive factor
25	474.5	10.1	625	1 KFHU	coagulation factor
26	471	10.0	275	1 KFHU	coagulation factor
27	469	10.0	810	2 S40007	trypsin (EC 3.4.21
28	468	10.0	810	2 I46260	plasma (EC 3.4.21
29	467.5	10.0	2616	2 A57096	model protein prec

30	466.5	10.0	452	1 A30351	coagulation factor
31	463	9.9	1420	2 A32869	apolipoprotein(a)
32	461	9.8	455	2 A61545	plasma (EC 3.4.21
33	460	9.8	275	2 S40005	trypsin (EC 3.4.21
34	458	9.8	562	1 UKHOT	c-plasminogen acti
35	455	9.7	274	2 S35339	trypsin (EC 3.4.21
36	454.5	9.7	267	2 S40006	trypsin (EC 3.4.21
37	453.5	9.7	560	1 JC4795	plasma hyaluronan-
38	444	9.5	228	1 TRW5Y	trypsin-like prote
39	441.5	9.4	1004	2 T30338	oviductin (EC 3.4.
40	440.5	9.4	4548	2 S00657	apoprotein(a) (EC
41	439	9.4	254	1 TRW3Y	trypsin-like prote
42	434	9.3	263	2 A21195	chymotrypsin (EC 3
43	432	9.2	242	2 S49489	trypsin (EC 3.4.21
44	432	9.2	263	1 KYRB	chymotrypsin (EC 3
45	430.5	9.2	247	1 TRDG	trypsin (EC 3.4.21
46	430	9.2	270	2 S56160	mast cell tryptase
47	429.5	9.2	459	2 JQ0419	coagulation factor
48	426	9.1	277	2 S35340	trypsin (EC 3.4.21
49	425.5	9.1	407	1 KFBQ	coagulation factor
50	425	9.1	265	2 T15451	hypothetical prote
51	424.5	9.1	231	1 TRPGR	trypsin (EC 3.4.21
52	423.5	9.0	558	1 JC5878	plasma hyaluronan-
53	422	9.0	231	2 S31778	trypsin (EC 3.4.21
54	421.5	9.0	247	1 B25852	trypsin (EC 3.4.21
55	420.5	9.0	247	2 A27547	trypsin (EC 3.4.21
56	420.5	9.0	304	2 S33496	trypsin (EC 3.4.21
57	420	9.0	263	2 A31299	chymotrypsin (EC 3
58	419.5	9.0	703	2 A47246	trypsin (EC 3.4.2
59	419.5	9.0	273	2 C1HURB	complement subcomp
60	419	8.9	246	2 JQ1472	trypsin (EC 3.4.21
61	418	8.9	245	2 KYBOB	chymotrypsin (EC 3
62	417.5	8.9	761	2 JC5759	brain-specific ser
63	417	8.9	246	2 JQ1471	trypsin (EC 3.4.21
64	417	8.9	276	2 A38654	mast cell proteinase
65	415.5	8.9	246	1 TRRT1	trypsin (EC 3.4.21
66	415.5	8.9	266	2 S54146	trypsin (EC 3.4.21
67	415.5	8.9	482	1 EXRT	coagulation factor
68	414	8.8	229	1 TRDS	trypsin (EC 3.4.21
69	414	8.8	246	2 B25528	trypsin (EC 3.4.21
70	413.5	8.8	456	1 KXBO	thrombin (EC 3.4.2
71	413.5	8.8	622	1 TBHU	trypsin (EC 3.4.21
72	413	8.8	242	2 S31776	trypsin (EC 3.4.21
73	413	8.8	242	2 S31775	trypsin (EC 3.4.2
74	413	8.8	274	2 JC4171	trypsin (EC 3.4.2
75	412	8.8	247	2 S05494	trypsin (EC 3.4.21
76	412	8.8	275	2 A32410	chymotrypsin (EC 3
77	409	8.7	263	2 S47537	u-plasminogen acti
78	409	8.7	432	1 S18932	trypsin (EC 3.4.21
79	408.5	8.7	247	2 S13813	trypsin (EC 3.4.21
80	408.5	8.7	231	2 S47538	acrosin (EC 3.4.21
81	408	8.7	259	2 I38363	trypsin (EC 3.4.21
82	408	8.7	275	2 B35863	trypsinase (EC 3.4.2
83	408	8.7	559	2 A35029	c-plasminogen acti
84	408	8.7	618	2 A35827	thrombin (EC 3.4.2
85	407.5	8.7	264	2 I38336	chymotrypsin-like
86	407	8.7	266	2 JC4850	trypsin-like prote
87	407	8.7	275	2 A35863	trypsinase (EC 3.4.2
88	407	8.7	559	2 A29941	c-plasminogen acti
89	406	8.7	245	1 KYBOA	chymotrypsin (EC 3
90	406	8.7	466	1 KFHU	coagulation factor
91	405.5	8.7	238	2 S31779	trypsin (EC 3.4.21
92	404.5	8.6	237	1 TRCV1	trypsin (EC 3.4.21
93	404.5	8.6	274	2 I47078	coagulation factor
94	404	8.6	247	1 A25852	trypsin (EC 3.4.21
95	403.5	8.6	250	2 S55493	serine proteinase
96	403	8.6	271	1 ELRRT	pancreatic elastase
97	403	8.6	275	2 C35863	trypsinase (EC 3.4.2
98	402.5	8.6	246	1 TRRT2	trypsin (EC 3.4.21
99	402.5	8.6	247	2 S12764	trypsin (EC 3.4.21
100	402.5	8.6	461	1 S18994	protein C (activat

## ALIGNMENTS

## RESULT 1

membrane-bound arginine-specific serine proteinase precursor - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C/Accession: J07731; J07775

R/Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Achnada,

J. Biochem. 130, 425-430, 2001

A/Title: Characterization of a membrane-bound arginine-specific serine protease from rat

A/Reference number: J07731; PMID:21421307; PMID:11530019

A/Accession: J07731

A/Molecule type: mRNA

A/Residues: 1-855 <R1S>

A/Cross-references: UNIPROT:Q9JUI7; DDBJ:AB049189

A/Experimental source: strain Male, 7-week-old

R/Satoml, S.; Yamazaki, Y.; Tsubuki, S.; Hicomi, Y.; Iwanaga, T.; Fushiki, T.

Biochem. Biophys. Res. Commun. 287, 995-1002, 2001

A/Title: A role for membrane-type serine protease (MT-Sp1) in intestinal epithelial tu

A/Reference number: J07775; PMID:11573963

A/Contents: Small intestine

A/Accession: J07775

A/Molecule type: mRNA

A/Residues: 1-855 <SAT>

A/Cross-references: DDBJ:AB037898

C/Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease

of specific proteins or peptides on the brushborder membranes. It also participates in

lial migration and/or cell loss.

C/Genetics:

A/Gene: mt-spl

A/Map position: basolateral cell surface

C/Superfamily: membrane-bound arginine-specific serine proteinase

C/Keywords: protein digestion

Query Match 82.6%; Score 3866; DB 2; Length 855;  
Best Local Similarity 80.8%; Pred. No. 6.1e-245;  
Matches 691; Conservative 79; Mismatches 85; Indels 0; Gaps 0;

```

QY 1 MGSRRARKGGGPPDRGAGLKYNRRHKNVGLGEGVFLLPNNVKKYKIKGPGWVLLA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MGNRRKAGGGSDPFAGLKYNRLNNMGFEBSVFLLPNNAKQYKGGPRWVWVA 60
QY 61 VLIGLLVLGIGLVWHLQYRDVVOVKVNGYRRTINENFVADYKSNSTEFYLSKV 120
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VVRSFLLSLMAGLLVHFRVNRVIRQVNGHRTINENFLDYKSNSTEFYLSKV 120
QY 121 KDAIKLLYGVPLGYPYHKESAVTAFSEGSVIAYVWSEFSIPOHLVEAEKRWAAEERV 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 KEALKMYSEVPVLPYHKKSTVAFSEGSVIAYVWSEFSIPHLBEVDRAVAVERV 180
QY 181 LPPARSLKFPVYVSVAFPPDSKTVORTDNGSGFGLHARGVLAFFTPGFPDPSYPA 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 LPPARSLKFPVYVSVAFPPDSKTVORTDNGSGFGLHARGVLAFFTPGFPDPSYPA 240
QY 241 HARCQWALRGDADSVLSLFRSFDIACDERSGDLVTVYVTLSPMEPHALVOLCGTTPS 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 HARCQWALRGDADSVLSLFRSFDIACDERSGDLVTVYVTLSPMEPHALVOLCGTTPS 300
QY 301 YNLTSSQVNVLLTTLTNTERRHGFEGATFFQLPKMSCGGLRLKAKOGTFNSPYRGHY 360
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 YNLTSSQVNVLLTTLTNTERRHGFEGATFFQLPKMSCGGLRLKAKOGTFNSPYRGHY 360
QY 361 PPNIDCWNIENPVNOMVYKRFKFFYLLERGVPAVGTCPKQVVEINERKCYGERSQFVTS 420
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 PPNIDCWNIENPVNOMVYKRFKFFYLLERGVPAVGTCPKQVVEINERKCYGERSQFVTS 420
QY 421 NSNKTIRFHSDDOSTDTGFLAEYLSYDSDPCPGQFTCTRGTCIRKELRCDDGADCTDH 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 NSNKTIRFHSDDOSTDTGFLAEYLSYDSDPCPGQFTCTRGTCIRKELRCDDGADCTDH 480
QY 481 SDEINSCDAGHQTCKNKKCPPLFWVCDSVNDGSDSGSCPAQTFRCSNGKCLSK 540

```

## RESULT 2

enteropeptidase (EC 3.4.21.9) precursor [validated] - pig

C/Species: Sus scrofa domestica (domestic pig)

C/Date: 07-Oct-1994 #sequence\_revision 09-Aug-1996 #text\_change 28-Apr-2003

C/Accession: A53663

R/Matsushima, W.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa, I

J. Biol. Chem. 269, 19976-19982, 1994

A/Title: Structural characterization of porcine enteropeptidase.

A/Reference number: A53663; PMID:94327548; PMID:8051081

A/Accession: A53663

A/Molecule type: mRNA

A/Residues: 1-1034 <MAT>

A/Cross-references: GB:D30799; NID:G505122; PIDN:BA06459.1; PID:G505123

A/Note: parts of this sequence, including the amino ends of three chains isolated from ct

ated below) or with amino-terminal myristoylation of the heavy chain.

C/Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)

ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involve

C/Function:

A/Description: cleaves activation peptide from trypsinogen to produce active trypsin

A/Pathway: intestinal digestive hydrolase cascade

C/Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding rep

C/Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen

F/22-38/Domain: transmembrane #status: predicted <TM>

F/52-117/Product: enteropeptidase mini chain #status: predicted <MCH>

F/118-799/Product: enteropeptidase heavy chain #status: predicted <HCH>

F/199-204/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F/357-519/Domain: MAM homology <MAM>

F/541-646/Domain: Clr/Cls repeat homology <CLR>

F/658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F/693-798/Domain: scavenger receptor cysteine-rich domain homology #status: atypical <SRCF

F/800-1034/Product: enteropeptidase light chain #status: predicted <LCH>

F/116-147/Domain: trypsin homology <TRY>

F/787-811, 825-841, 925-992, 956-971, 982-1010/Disulfide bonds: #status: predicted

F/840, 891, 986/Active site: His, Asp, Ser #status: predicted

Query Match 15.5%; Score 724; DB 1; Length 1034;

Best Local Similarity 23.8%; Pred. No. 2e-39;

Matches 258; Conservative 135; Mismatches 330; Indels 360; Gaps 41;

```

QY 55 WVTLLAVLIGLLVLGIGLVWHLQYRDVYVK--VKNGYRIT-----NEKPVAYEN 107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      21 FTALFALIMVLCAGLIIVSWLTIKGEKDALGKSHARGMTKITSVTYVNIQDGL-- 78
Qy      108 SNSTEFVSLAKVDALKLYSGVFLGPYHKEAIVTAFSGSVAYV-----W-SFES 160
Db      79 --SVDFKVLADIDQIMCEIFQSSNLIKNEY-KNSRVLOFENGSGVIVFDLLFAQVSDEN 135
Qy      161 IPOHLVEAEKVMKEERVVM----- 180
Db      136 IKEELIQTIEANKSSQVAFHIDVNSIDITSELENYSTTSPSTSDKLTSSPPATPGNV 195
Qy      181 ----LP--PRASLSKSFV-----TSVAFPTDSKTORTDNCSFGIARHAGVELMRF 229
Db      196 SIECLPGSRPCADLAKIAVLDFCGELNCPDGSDESKICATACD-----GKFLIES 249
Qy      230 TPGPDSPYP-----AHARCOALRGDADSVLSLTFRSGLASCDE-----GSDLV-- 276
Db      250 SGSPFAAQYPRKLEASVVCWMIIRVNOGLSIELNFSYENTYSMDVNIYEGSGSKILRA 309
Qy      277 -----TVNYTILS----- 283
Db      310 SLMTMNPETIRIFSNQVTVTLFIESDENDYIGFNATYTAFTASTELANDEKINCNEDEGFC 369
Qy      284 -----PM 285
Db      370 FWIDLNDNDEMERICGTPPTPGPNPDHTFGNASGYIISPTGPGRGQERVGILSLPL 429
Qy      286 EPHALVOLC-----GTYPSYV----- 302
Db      430 EP-TLEPCLSPWYMGENVYKLSINISNDONNEKIIPOKEGNGEMWNNGOVTLNETV 488
Qy      303 ----LTFHSSQVLLITLTNTERHRRGR-----EATPF-----QLRMSSCG--RLR 345
Db      489 EFKVAFNFKQFLSDIALDILSLTYGICNVSLPEPLVTPSPPEL--TDCCGPEFLW 546
Qy      346 KAOGFNSPYRPGHYPNIDCTWNIENPNOHVKVRFKFFYLLEBPGAGTCPPXUYEI- 404
Db      547 EPNFTTFSMNPNNYPRQAFCCWMLNAGKKNIOHFEFLENIA-----DVVEIR 598
Qy      405 NGEK-----YCGERSQFVYTSNSKITYRFHSDQSYDTGFLAEVLS--YDSQPC 453
Db      599 DGEEDSDLLAVYTGPGFVEDVFFSTNNMTVLFTINDALTGKGFANFTTYGHLIPRCP 658
Qy      454 -PGFTCTGRGCIKRELACDGMADCTDHSDELNCSGADHQFTCKNKKCKPLFWCDSVN 512
Db      659 KEDNQCNGEBCVLLVNLCDGSHCKDQSDDEAHCV-----RP-----LN 697
Qy      513 DCGNDSDEGSCSPAOTFRCS--NGKCLSKSQCCNGKDC--GDGSDASCP----- 560
Db      698 GTANNSG-----LVQFRIGSIWHTACAEWMTTQSDVQQLGLGTGNSMFPFSSCG 750
Qy      561 ----KVVVVTCTKHTYRCLNGICLSKGNPECDGKED-----CSDGSDKDCDGLRSFT 610
Db      751 GPFVLTNTPA-----NGSLILITASEQC--FEDSLILILQCNHKS-----CGKQOVA 793
Qy      611 RQA--RVVVGTDADGEGWPMQVSLHAGQHICGASLISPMWIVSAAHCIYDDGFRSD 668
Db      794 QEVSPKTYIGANDSRGAMPVVALYVNO--LTCGSLVSRMLVSAALCVYG---RULE 848
Qy      669 PTOMTAFGLIDOSGASVGOERLKRILSHPFENDTTPDYALALELEKPAEYSSWVR 728
Db      849 PSKMKAILGLHMTSLTSPQIVTRLIDELVINPHNRRKOSDIAMHLEKFAVYTDIQ 908
Qy      729 PICLPDASHVPPAGKAIWVGKHTQYGGTGLILQKEIRIVINOTTCENILPO-QITPR 787
Db      909 PICLPBENQVPPPGICISIAWGKVIYQSGPADILQEDADVPLLSNEKQOQMPENITEN 968
Qy      788 MMCGVFLSGGVDSCGDGGPLSSVEADGRIFGAGVSWMGDCGRNKGYYTTRPLPRD 847
Db      969 MMCGYEEGGIIDS CGDGGPLMCLL--NNRMILAGVTSFGYQCALPNRPGVYARVFKTE 1027
Qy      848 WIK 850
Db      1028 WIQ 1030

```

---

```

RESULT 3
A:3090
enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine
N:Alternate names: enterokinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A43090; A4874; A61436
R:Kilamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
A:Reference number: A43090; MUID:94329561; PMID:8052624
A:Accession: A43090
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1035 <NIT>
A:Cross-references: UNIPROT:P96072; GB:U09859; NID:G746410; PIDN:AAB40026.1; PID:G746411
J: Biol. Chem. 268, 23311-23317, 1993
R:Valle, E.R.; Rehmetulla, A.; Racie, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.; I
J: Biol. Chem. 268, 23311-23317, 1993
A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of bc
A:Reference number: A4874; MUID:94043122; PMID:8226855
A:Accession: A4874
A:Molecule type: mRNA
A:Residues: 801-1035 <LAV>
A:Cross-references: GB:L19663; NID:G416131; PIDN:AAA16035.1; PID:G416132
A:Note: parts of this sequence, including the amino end of the mature protein, were confi
R:Light, A.; Janska, H.
J: Protein Chem. 10, 475-480, 1991
A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.
A:Reference number: A61436; MUID:92189715; PMID:11799406
A:Accession: A61436
A:Molecule type: protein
A:Residues: 801-807, 'Y', 809-827 <LIG>
C:Comment: The mechanism of association with the membrane of the intestinal brush border
embrane attachment using a signal-anchor sequence.
C:Comment: Conversion from membrane-bound to soluble forms may involve further processing
C:Complex: mature enteropeptidase is variously reported to contain two (heavy and light)
hide linked
C:Function:
A:Description: cleaves propeptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Ctr/Cis repeat homology; LDL receptor ligand-binding rep
C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein.
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-800/Product: enteropeptidase heavy chain #status predicted <HCH>
F:119-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:356-520/Domain: MAM homology <MAM>
F:542-647/Domain: Ctr/Cis repeat homology <CTR>
F:659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:694-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRCT
F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>
F:801-1030/Domain: trypsin homology <TRY>
F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Binding
F:786-912,926-842,926-993,957-972,983-1011/Disulfide bonds: #status predicted
F:841,892,987/Active site: His, Asp, Ser #status predicted

```

---

```

Query Match 15.2%; Score 709.5; DB 1; Length 1035;
Best Local Similarity 23.1%; Pred. No. 1,7e-38;
Matches 251; Conservative 150; Mismatches 320; Indels 365; Gaps 44;

```

---

```

Qy      55 WVTLAVLVIGLLVGLIGLVLMHLYRDV---RVQKVGKNGKRNITENFVDVYNSNST 111
Db      21 FAVLFLVILVLCAGLIIVSWLISIQSVKDAFGKSHARGTILKISGATVNPHLQDGLSV 80
Qy      112 EPLVSLAKVDALKLYSGVFLGPYHKEAIVTAFSGSVAYV-----W-SFES1PQH 164
Db      81 DFKVLADIDQIMDDIFQSSNLIKNEY-KNSRVLOFENGSIIVFDLLFDQVSDKNKKEE 139
Qy      165 LVEAEKVMKEERVVM-----LPPRSLSKSFVVTYSV-----V 197

```

Db 140 LIQIEANKSSQVLTVEHIDINSIDITASLENFSTISPAITSEK--LTTSIPLATPGNVS I 197  
Qy 198 AFPTDSKTVORTOQNSGFGHARGVEIM-----RFTTPEGF 233  
Db 198 ECPFDSKL-----CADALKCLAIIDFCDGELNCPDSDENKTCATACDGFLLTGS 249  
Qy 234 PDS-----PYPAH--ARCGMALRGDADSVLSITFRSFD----- 264  
Db 250 SGSEFALHYPKFSNNTSVAVCGWILIRVNOGLSIQNFYFNTYADVNLNIEYEGMGSSKILR 309  
Qy 265 -----LASCDERGSDDL--TVNTLSLPMEPHALVOL----- 293  
Db 310 ASLWSNPGILRIIFSNVOTATFLLQSDS--SDYIGFVVTYAFNSKELNNEYKINCNPED 367  
Qy 294 --C-----GTYPPSYNLPF-HS----- 307  
Db 368 GFCTWIDLDNDNEMERTQSGTFFPSTGPTFDHIFGMSGYISTPTGPGRRRVGLLT 427  
Qy 308 -----SONVLLITLITNTERRHGFEATFPQ----- 333  
Db 428 LPLDPTBQACLSPWYMYGENVYKLSINISSQ--NMEXTIRPKEGNGVQNNVQVLT 484  
Qy 334 -----LPMSSCGG 342  
Db 485 LNETVERKVSFYGKNOILSDIALDIDSLTYGICNVSVPEPILVTPPPELP--TDCCG 542  
Qy 343 --RLRKAQGTENSPYYPGHYPNNIDCTWNIEVPNNQHVKRFKFFYLLEPGVPAGTCKPD 400  
Db 543 PHDMEWNTTFTSINFPNSYNQNAFCIWNLNAGKNIOQLHFQDFLENTA-----D 594  
Qy 401 YVEI-NGEKYCGERSQFV-----VTSNSKITVRPHSDQSYTDTGFLAEYLS-- 446  
Db 595 VVEIRDS--GDSLSFLAVYTGPGVNDVSTNRMVTVEFITNMLAKQCFKANFTTGY 651  
Qy 447 -YDSSDC-EGQFCRCRGRCIRKELRCDGMADCTDHSDELNC-----SCDAGH--QFTC 496  
Db 652 GLGIPCEKEDNFQCKDGECEPIVNLCDGFPHCKDGDDEAHCVRLFNGTDSGLVQFRI 711  
Qy 497 KNKFCRKLFWVCDVNDGSDNSDEQGSCEPAQTFRCNGSKCLSKSQCGCKGDKDGDGSDS 556  
Db 712 QS-----IMHV--ACAEN-----WTQIISDDVC-----QLLG--LGTGNS 743  
Qy 557 ASC-----PKRVNVTCTKHTYRCLNGCLSKNGPECDGKEDCSGSDDEKDCD--CGLR 607  
Db 744 VPTSTGGGPGVNNTPAP-----NGSLITLTPSQOC-----LEDSSLILQCNYSKCGKK 791  
Qy 608 SFTFQA--RVVGGDADGEMPMQVSLHALGCGHIICGASLSPMWVSAHAYCIDDDGFR 665  
Db 792 LVTOESVPKIVGSDSREBAMPVVALYFDQO--QVCASLVSRLMVSAAHCVYG---R 846  
Qy 666 YSDPTQWTAFLGLHDQSRAPVQOERRLKRIISHPEFNDFTPDYDIALLELEKPAEYSS 725  
Db 847 NMESKKAIVGLHMASNLTSFQLETRLLIDIVINPHNKRKKNDIAMMHELEKNAVYTD 906  
Qy 726 MVRPICLPDASHVFPAGKAIWVTGHTQYGGTALILQKEIRVINQTTCENLIPQ-QI 784  
Db 907 YIQPICLPEENQVPPRISCISIAQMGALIIYQGSTADVLQEDAVPLINSEKCOQOMPEXNI 966  
Qy 785 TPRMMVCGLFSLGVDSCGDSGGLSVEADGRIFGAGVWSKGGCGCKRKPQVYTRLLPL 844  
Db 967 TENNVCAGYEAGVDSQGDGSGPLMCOE--NNRWLLAGVTSFGYQCALPMRPGVYAAVPR 1025  
Qy 845 FRDWIK 850  
Db 1026 FTEWIO 1031

## RESULT 4

A56318  
enteropeptidase (BC 3.4.21.9) precursor [validated] - human  
N:Alternate names: enterokinase  
C:Species: Homo sapiens (man)  
C:Date: 19-May-1995 #sequence\_revision 09-Aug-1996 #text\_change 09-Jul-2004

C:Accession: A56318; B43090  
R:Klitamoto, Y.; Veille, R.A.; Denis-Keller, H.; Sadler, J.E.  
Biochemistry 34, 4562-4568, 1995  
A:Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic  
A:Reference number: A56318; MIM:195234679; PMID:7718557  
A:Accession: A56318  
A:Molecule type: mRNA  
A:Residues: 1-1019 <KIT>  
A:Cross-references: UNIPROT:P98073; GB:U09860; NID:g746412; PID:AA50138.1; PID:g746413  
R:Klitamoto, Y.; Yuan, X.; Wu, O.; McCourt, D.W.; Sadler, J.E.  
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994  
A:Title: Enterokinase, the Initiator of intestinal digestion, is a mosaic protease compo  
A:Reference number: A43090; MIM:194329561; PMID:8052624  
A:Accession: B43090  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 749-1019 <KIT>  
A:Cross-references: GB:U09860  
C:Comment: The mechanism of association with the membrane of the intestinal brush border  
ated below) or with amino-terminal myristoylation of the heavy chain.  
C:Genetics:  
A:Gene: GDB:PRS57  
A:Cross-references: GDB:384083; OMIM:226200  
A:Map position: 21q21-21q21  
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)  
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involve  
ducts.  
C:Function:  
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin  
A:Pathway: intestinal digestive hydrolase cascade  
C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding rep  
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen  
F:1-784/Product: enteropeptidase heavy chain #status predicted <HCH>  
F:22-38/Domain: transmembrane #status predicted <TMM>  
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
F:342-504/Domain: MM homology <MM>  
F:546-631/Domain: Clr/Cls repeat homology <CLR>  
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRCF  
F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>  
F:785-1014/Domain: trypsin homology <TRY>  
F:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site: c  
F:772-996,910-826,910-977,941-956,967-995/Disulfide bonds: #status predicted  
F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 14.5%; Score 679.5; DB 1; Length 1019;  
Best Local Similarity 32.5%; Pred. No. 1.6e-36;  
Matches 180; Conservative 80; Mismatches 202; Indels 91; Gaps 20;

Qy 333 QLRPMSSCGG--RLRKAQGTENSPYYPGHYPNNIDCTWNIEVPNNQHVKRFKFFYLLEP 390  
Db 519 ELF--TDCGPFLEWBNPTFTSTNFPNSYPLAFVWILNAGKNIOQLHFQDFLENT 575  
Qy 391 GVPAFCPKDYVEI-NGEK-----YGERSQFVYTSNSKITVRPHSDQSYTDTGFL 441  
Db 576 -----INDVVEIRIGEENDSLLAIVYTGPGVNDVSTNRMVTVEFITNMLAKQCFKANFTTGY 651  
Qy 442 AEVLS--YDSSDCPG-OFTCRGRCIRKELRCDGMADCTDHSDELNCSADAGHOF--- 494  
Db 629 ANFTTGHGIPFCADHFQCKNGECVPLVNLCDHGLHCEBDSDDADVC---RFPNG 663  
Qy 495 TCRN---KFCRKLFWVCDVNDGSDNSDEQGSCEPAQTFRCNGSKCLSKSQCGCKGDKD 550  
Db 684 TNNNGLVRRFRIQSIW--HTCAEN-----WTQIISDDVC-----QLL 719  
Qy 551 GDGSDASCPKXNVVVTCTKHTYRCLNGCLSKNGPECDG-----KEDSDGS-----D 598  
Db 720 GLGSGNSKRIFFST-----DGGPFVXKNTAPDHLITPTSQOCLQDLSRLQCN 768  
Qy 599 EKPCDGLSFTQARVAVGTADGEMPMQVSLHALGCGHIICGASLSPMWVSAHAYCIDDD 658  
Db 769 HKSCKKLAADITTPKIVGSDSREBAMPVVALYFDQO--QVCASLVSRLMVSAAHCVY 827



```

OY 659 IDDR6FRISDPQWMTAFGLHDQSORBAGVERLKLK1ISHPFNDSTPYDIALLELE 718
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 828 YG---RNLBPBKWAI1GLHKMSKULTSPQVPR1DE1VINPHYNRRRKND1AMHLE 883
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 719 KPABYSWVRP1CELPDASHVPFAGKAIWYTGWGH1QYGGTAL1J0KEIRVYN0TGEN 778
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 864 FKAVNTDY1QPC1PEBNQVPPRGNC1S1AGM1VYGGT1N1IQEADV1LS1ERK1CQ 943
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 779 LLPQ-Q1TPRMVCVFLSGVDSCQSDSGSLSSVEADGR1FGAGVSWGDGCRANKRG 837
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 944 QMPEXN1TENN1CAGYEEGG1DSCQSDSGGLMCGE-NNRWPLAGVTSFGYKCALPNRPG 1002
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 838 VYTR1PLFRDW1K 850
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1003 VYARVSRFTEW1Q 1015
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 5

low-density lipoprotein receptor-related protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 16-Jul-1999 #sequence 16-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: J03035  
R:Tomlita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.  
J. Biochem. 124, 784-789, 1998  
A:Title: A novel low-density lipoprotein receptor-related protein with type II membrane  
A:Reference number: J03035; MUID:96429596; PMID:9756624  
A:Accession: J03035  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1113 <TOM>  
A:Cross-references: UNIPROT:Q9Z319; DBJ:AB013874; NID:g3669144; PION:BA03371.1; PID:g3669144  
C:Superfamily: mouse low-density lipoprotein receptor-related protein; LDL receptor ligand  
F:337-372/Domaln: LDL receptor ligand-Binding repeat homology <LDL1>  
F:374-408/Domaln: LDL receptor ligand-Binding repeat homology <LDL2>  
F:410-445/Domaln: LDL receptor ligand-Binding repeat homology <LDL3>  
F:447-482/Domaln: LDL receptor ligand-Binding repeat homology <LDL4>  
F:646-682/Domaln: LDL receptor ligand-Binding repeat homology <LDL5>  
F:684-720/Domaln: LDL receptor ligand-Binding repeat homology <LDL6>  
F:723-757/Domaln: LDL receptor ligand-Binding repeat homology <LDL7>  
F:863-1097/Domaln: trypanin homology <TRY>

Query Match	14.1%	Score 661.5	DB 2	Length 1113
Best Local Similarity	32.7%	Pred. No.2,6e-35		
Matches 158	Conservative 67	Mismatches 187	Indels 71	Gaps 16
Oy	428	RPHSDOSTYDGFLEAYLVSYDSDPCPGQFTCRGCRKRELRCDGMADCTDHSDELNCS	487	
Db	627	QPESSSNQTCILP--NEDVEBCSPSHFKRSRCVLSGRRCDDGADCDDBDEBNC	683	
Oy	488	CDAGHQFTCK-NKFKCPLFWWCDSVNDGSDNDEGCS-CPAQFTFGCSNGKCLSKSQCN	545	
Db	684	CKERALMCEPFGKQCLKTKHTLICDGFPPDCPDSDMEKNCFCODNELECANHECVPRDIWCD	743	
Oy	546	GKDDCGDSDSEASCNKVN-----VVTCTKTYTCLNG-----LC-----	579	
Db	744	GWVDDSDSDDEGCVTLTSKNGNSSLLTVHKSAKEHVCADGMRETLSQLACKQMGCEP	803	
Oy	580	-LSKNGPECDGKE-----DCSDGSDSEKD-----C--DCGLRSFT	610	
Db	804	SVTKLIPQEGEQMLRLYPRNMENLNGSTLQELLYVRHS CSPKSEISILCSKQDCGRRAA	863	
Oy	611	R-QARVVGCTDADBESEWPMQVSYHALGGHGTGASLISPMNLVSAACHTIDRGRISDP	669	
Db	864	RNNKKIILIGRTSRSPRMPWQCSLOEBSGHTCGCVLAKKWLTVAAHCF-BGR-----EDA	918	
Oy	670	TQMTAFLLGLHDQSGSAPGVQERRLKRIISHPFVDFPDYDIALLELEKPAEYSWVR	729	
Db	919	DVKKVVFGIINNLDHSG-FMQTRPYKTIILHRRYSRAVVDIDISVELSDDINEISTYRP	977	
Oy	730	ICLPDASHVPFAGAKIWTGWTGYGTGALILQKGEIRVINTQTCENLLPQ-ITPRM	788	
Db	978	VCLPPEPEYLEDPTCYITGWGH--MGNMMPKQDEGVRIILPEQCQSYDMKTIITRM	1035	

QY	799	MCVGEFLSGGVNDCQGDSCGPLSSVEADQRIFAGAGVSWGDCACGR-NKPGVYTRLPLEFRD	847
DB	1036	ICAGIESSTVDSCMGDSCGPLVCERPGQWTLFGILTSMGVSCFSKVLGPGVYSHNVYFVG	1095
QY	848	WIK	850
DB	1096	WIE	1098

## RESULT

polypeptide - African clawed frog  
 C.Species: *Xenopus laevis* (African clawed frog)  
 C.Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
 C.Accession: T30337  
 R.Yang, J.C.; Lindsey, L.L.; Hedrick, J.L.  
 submitted to the EMBL data library, March 1998  
 A.Description: cDNA cloning of oocyte, a chymotrypsin-like protease released from *Xenopus* oocytes  
 A.Reference number: Z20829  
 A.Accession: T30337  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: mRNA  
 A.Residues: 1-1524 <YAN>  
 A.Cross-references: UNIPROT:Q91674; EMBL:U081290; NID:G2981640; PID:G2981641; PIDN:ACC2477  
 C.Superfamily: trypsin related polypeptide; trypsin homology

Query Match	12.3%	Score 576.5;	DB 2;	Length 1524;
Best Local Similarity	26.2%;	Pred. No. 1.3e-29;		
Matches 167; Conservative	89;	Mismatches 205;	Indels 177;	Gaps 23;

```

Qy 238 YAHARCCWALGDADSVSLTFRSFDL---ASCDSRGSVLVYNNLTS-PMHEHALVOL 293
Db 333 YSINSRCRMMLLVQKAKTIEIRFLQJLDBDHAC-----TFDYLSFTVNEKMKIRK 383
Qy 294 CGTYPSPYMLTFHSSONVLLITLINTERRHPGEATFPQLP-RMSSCG-GRUKKAGT 350
Db 384 CGSTIPSPILV--RKNKVTVPFFSDGTFTGRGFEIOFLAIPFKASACGSAKILKKKM 440
Qy 351 FNSPYRPGHYPPENIDCTWNIEVPNNQHKYRFEFFYLEEGVPAQ-TCPPDYVI-NG-- 406
Db 441 IYSPNYPDPYPLUKTSMWILEAENHIVKLKFFDF---NVEYGHCIYDAVEYVQAE 495
Qy 407 -----EKYGEKRSQFVVTNSNKTIVRFHSDOSYTDTGFLAEYLSYDSBPCQPTCRT 461
Db 496 EKQLLARLQGYLLPLPFISSPENTMLRFRKTMENSVGFVKF----- 538
Qy 462 GCIRKELKCDGMADCTDHSDELNCSCDAGHOTCKRK-FCKPLFWCDSVNDGDNDSDE 520
Db 539 -----SVPKPKOFSLPV----- 551
Qy 521 QGSCCPAQFRCNSGKCLSKSQCGCKGDCGSDSEASCKPVVVVCTKHATYRCLANGCL 580
Db 552 -----DB-----PTIEML----- 560
Qy 581 SKGNPECDCKEDCSDGDEKDCDGLRSFTRQ---ARVCGTDADEGEMVQUSLHALQ 637
Db 561 ---HPAIALD-----VCGMAPMTPKMWLPRIIVGEEBASPNSWPMQVQIFELRT 606
Qy 638 GHICGASLSPMMLVSAACHYIDDRGFRYSDPOMTAFLLJHPOSQASAPGVQERRAKRI 697
Db 607 FH-CERAAITSPQWILFNAHC-----IRAAEPYWIYIADNHRMLNEST-EQIRNKTI 658
Qy 698 ISHPFENDEFDYIALLELEKPAEYSSMWRPICLPDASHVPFPAKAIWYTGWGHTOYG 757
Db 659 RIHDYNNSETYNDIALLLYLEEPLDLNDFVRPCLPEBEEVLTPASVCVVTGMENTABDG 718
Qy 758 TQALLQKEIIVINOTCE-NLLPQOITPRMKCVGFLSG-GVDSCCGSDSGGPLSYEAD 815
Db 719 QPALGIQQLQLOLPLBDSITICNTSYSGELTDHMLCAGFPSSKEKDDACGSDGGGPLVCONEK 778
Qy 816 GRIFGAGVSWGDGCAGRNKPQVYTRPLFRDMIKENT 853

```

Db 779 BQFSIYGLVMSGCGRVSKPGVYTKVRLFFTWI-QNT 815

RESULT 7

154763

Ra-reactive factor (EC 3.4.21.-) 1 precursor - human

N/Alternate names: mannose binding protein-associated serine proteinase 1 (MASP-1)

C/Species: Homo sapiens (man)

C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004

C/Accession: 154763; JN0883

R/Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.

Int. Immunol. 6, 665-669, 1994

A/Title: Molecular characterization of a novel serine protease involved in activation of

A/Reference number: 154763; MUID:94289349; PMID:8018603

A/Accession: 154763

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-699 <S>

A/Cross-references: UNIPROT:P48740; GB:D28593; NID:g790963; PID:BAA05928.1; PID:g471128

R/Takada, F.; Takayama, Y.; Hatsuue, H.; Kawakami, M.

Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993

A/Title: A new member of the C1s family of complement proteins found in a bactericidal

A/Reference number: JN0883; MUID:9405062; PMID:8240317

A/Accession: JN0883

A/Molecule type: mRNA

A/Residues: 1-234,'E',236-284,'G',286-498,'K',500-542,'K',544-642,'S',644-699 <TR>

A/Cross-references: DDBJ:D17525; NID:9439712; PID:BAA04477.1; PID:g439713

A/Experimental source: liver

C/Comment: This is a serum bactericidal factor that activates complement C4 and C2 compo

C/Genetics:

A/Genes: GDB:MSP1, GDB:CRARF, CRARF1, PRSS5, MASP

A/Cross-references: GDB:361104; GDB:330954; OMIM:600521

A/Map position: 3q27-3q28

C/Superfamily: complement-activating serine proteases C1r/C1s/MASP; C1r/C1s repeat homol

C/Keywords: beta-hydroxyasparagine; complement pathway; duplication; glycoprotein; hydroc

F:1-17/Domain: signal sequence #status predicted <IG>

F:18-448,449-699/Product: Ra-reactive factor #status predicted <MAT>

F:19-135/Domain: C1r/C1s repeat homology <C1r1>

F:143-181/Domain: BGF homology <BGF>

F:185-294/Domain: C1r/C1s repeat homology <C1r2>

F:301-362/Domain: complement factor H repeat homology <FHL>

F:367-432/Domain: complement factor H repeat homology <FHL>

F:449-691/Domain: trypsin homology <TRY>

F:49,178,407/Binding site: carboxylate (Asn) (covalent) #status predicted

F:79-91,143-157,153-161,168-181,185-212,242-260,301-349,329-362,367-414,397-432,436-572,

F:159/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

F:448-449/Cleavage site: Arg-Ile (autolytic) #status predicted

F:490,552,646/Active site: His, Asp, Ser #status predicted

Query Match 12.2%; Score 570; DB 1; Length 699;

Best Local Similarity 25.3%; Pred. No. 1.5e-29;

Matches 193; Conservative 111; Mismatches 268; Indels 192; Gaps 32;

QY 203 SKTYQRFQDMSCSGFLARQVELMRFTTPGPDPSPVNAHRCQALRGDADSVLSTFRS 262

Db 15 SKSAHATVEILANNFG-----QIQSPGPPDS-YPSDEVTWNITVPDGFRIKLYFMH 64

QY 263 FDLASCDERSDLYTVNTLSPMEPHALVOLCG-----TTPPSY-NLTFH 306

Db 65 FNLSSVLCYDVYKV-----ETEDQVLATFCGHEFTTDQTQPGQEVVLSFGSFMSTTFR 119

QY 307 SSONVLLITLITTERHPGFEATFPOLP----- 335

Db 120 SD-----FSNEERFTGFDHAYMAVDVDECKEREDELSCDHYCHNYIGVYCSGRF 170

QY 336 -----RMSGCGRLKKAQGTENSPPYRGHYEPNDICTMNEVPPNNQHYKVAEPK 364

Db 171 GYLITMDNRFCRCVCSNMLFTQRTGVITSPDPFPPYKSSCYLTTELEEGFPMVNLQFED 230

QY 385 FYLLLE--PGVPAGTCPPDYVEIN-GEK-----YCGERSQFVNTSNKLTIVRFHSDQSYTD 437

Db 231 IFDIQDHPRP---CPDYIKIKKGPVKVLGPFCEKXAPPEPISQSHSVLLIFHSDNAEN 267

QY 438 TGFALVLYSDSDPCP-----GQFCTRGCIKREKCDGMADCTHSDLNSCDA 490

Db 288 RGMRLSYRA--AGNECEPIOPRVHGKIBPSQAKYFPK-----DQVLVSCDT 331

QY 491 GHQP-----TCKKKFCKPLFW-----VCSYVNOCCGNSDQGCGSCPAP-----TFR 531

Db 332 GYVVLKDNVEMDFQIECLYDGTWNSKIPCTKVD-----CRAPBLEHGLITFS 361

QY 532 CSNGKCLSKSQCGNGDCCGSDGSDENASCPRVNVVCTKHTYRC-----LNGCLSKGNP 585

Db 382 TRNNLTYYSEI---KYSCQ-----PYKMLNNNTGYTCSAGQWMNKV-LGRSLP 430

QY 586 ECDGKEDCSDGSDKDCDGLRSFTRO--ARVVGTDADGEMPMVWSLHALGQHI CGA 643

Db 431 TCLPV-----CGLPFKSRKLMARIFNGRPAPKGTTPWJAMLSHLNGQPFCCG 477

QY 644 SLSPMWVLSAANCY---SD--DRGRVSD---PLOTWTFGLGHDSQSAAPVQRRRLK 695

Db 478 SLTSSMWVYLAACHQSDPGDPTLRDSDLSPSDFKIILGNH-WRLSDENEQHLGVK 536

QY 696 RLISHPFNDFPDYDIALLEKPAEYSSMVRPCLPDASHVFPAGKA1WYTGWHTQY 755

Db 537 HTLHQQYDPTNFENDVALVELLESVLAFAVWPICLBGPQ--QSGANVIVSGWG-KQF 593

QY 756 GGTGALLDKGEIRVYNQTTCEML--LPQQTTPRMVCYGLSGVDSQCGDSGGLPSV 812

Db 594 LQRFPELTMEIEIPIDYHSTQCAVAPLKKKTVRDMICGEEKGGKADACGDSGPMVTL 653

QY 813 EAD-GRIFGAGVYSGDGCAGRRKPGVYTRLPFRMWEKNTGV 855

Db 654 NRERGVYLVGVYSGWDGCKKDRYGVYGIYHNKMKIQRVTVG 697

RESULT 8

KOMSP1

Plasma kallikrein (EC 3.4.21.34) precursor - mouse

C/Species: Mus musculus (house mouse)

C/Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 09-Jul-2004

C/Accession: A36557

R/Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochement

DNA Cell Biol. 9, 737-748, 1990

A/Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compari

A/Reference number: A36557; MUID:91090844; PMID:2264328

A/Accession: A36557

A/Molecule type: mRNA

A/Residues: 1-638 <SEI>

A/Cross-references: UNIPROT:P26262; GB:M58588; NID:9200358; PID:AAA63393.1; PID:g200359

A/Note: part of this sequence, including the amino ends of both the heavy and light chain

C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w

are linked by one or more disulfide bonds.

C/Superfamily: coagulation factor XI; trypsin homology

C/Keywords: blood coagulation; fibrinolysis; glycoprotein; hydrolase; inflan

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-330/Product: plasma kallikrein heavy chain #status experimental <HCH>

F:20-330/Domain: apple repeat <AP1>

F:110-109/Domain: apple repeat <AP2>

F:200-289/Domain: apple repeat <AP3>

F:291-380/Domain: apple repeat <AP4>

F:391-638/Domain: plasma kallikrein light chain #status experimental <LCH>

F:391-621/Domain: trypsin homology <TRY>

F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-

F:127,215,308,396,494/Binding site: carboxylate (Asn) (covalent) #status predicted

F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 11.3%; Score 528.5; DB 1; Length 638;

Best Local Similarity 27.2%; Pred. No. 6.8e-27;

Matches 184; Conservative 85; Mismatches 232; Indels 175; Gaps 32;

QY 271 RGSIDLTVVNTLSPMEPHALVOLCGTVPSPYVLTFFSSONVLLITLITTERHPGPF--E 328

Db 32 RGSIDLAAITTP-----DAQVCSQKMTNPRCLLFSF-----LAVTPRKETNKRFGCFMKE 81

```

Qy 329 ATFFOLPRMSSCG---GRLRKAQGTENSPYRGHYP-----NIDCTWNIE-----V 372
      :|||:
Db 82 SITGTLPRIRHTGALISGSHLKQCGHQSACHRDYIKGLDMRGSNFNISKNTNIEBQKLC 141
Qy 373 PNNQHVKKRFR---FYLLF-----PGVPACT----- 396
      :|||:
Db 142 TNNHFCQF--FYVAISAFRPEYRKXCLKLKHASGTPITSKADNLVSGFSLKSCALSEIG 200
Qy 397 CPKQ-----YVEINGEKCYGBERS---QFVNTSNKLTVAFHDSQSTVDG---FLAE 443
      :|||:
Db 201 CPMDIFQHSADLNVSGVITPDPAVCTICTFPHNCLFTFYTNWETESQRNVCFUKT 260
Qy 444 YLSYDSDPGCPQ-----FTCRTRG---CIRKELRCDMADCTHSDLELNCSDAGH 492
      :|||:
Db 261 SKSGRPSPIRPOENALISGYSLLTCKRKTRPECHSKL-----YSGVDPEGEELNATFVQGA 315
Qy 493 ---OFTCKNKFCKPLFWVCDVSV-NDCGNDSBQGCSCPAQTFRCGNGKLSKQCGNGKD 548
      :|||:
Db 316 DVCQETC-TKTRICQCFIYSLLPQDC---KEEGCKC---SLRLST----- 353
Qy 549 DCGGSDGASCPKXNVVTCGHTYRCLNG-----LCLSKNPREDDGKEDSDGSDKDC 602
      :|||:
Db 354 ---DQSP-----TRITYGQSGSGYSRLCLVDSPOCTTKIN----- 388
Qy 603 DCGLRSPTRQARVVGTDADGEMPMQVSLHA--LGQGHICGASLISPMVLVSAHACYID 660
      :|||:
Db 389 -----ARIGGTINASIGEMPMQVSLQVKSQHLCCGSLIGQWLTAAHCF-- 436
Qy 661 DRGFRYSPTQWTAFLGLHDQSQ--RSAPGVQERLKRILSHPPFNDFPYDIALLELE 718
      :|||:
Db 437 -DGIPY--PDVWRITYGILSLSEITKEPR---SSRIKELIHQEKVSEGVNDIALIKYQ 490
Qy 719 KPAVSSNVPRICLPDASHVFPACKALVWGMGHQVYCGTALLQKEILVINQTCEN 778
      :|||:
Db 491 TPLNLTBERQKPLCPKSKADNTTITYNCWVTGMYTKEGSETONILQKATIPLVNPEBQK 550
Qy 779 LLPQO--ITPRMCMCVFLSGVDSCQDSCGSLSVLEADGRIFGAGVWMDGCGAGRNKPG 837
      :|||:
Db 551 KYRIVYVINKQMICAGYKGGTACKDGBSGPL-VCKHSGRWQVIGITSWGSGCKRQKPG 609
Qy 838 VYTRLPLEFDMIKENT 853
      :|||:
Db 610 VYTKVSEYMDWILEKT 625

RESULT 9
KQRTPL
Plasma kallikrein (EC 3.4.21.34) precursor - rat
N:Alternate names: Fletcher factor; kininogenin; serum kallikrein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 09-Jul-2004
C:Accession: A39180; A33320; S06851; I53041; S06852
R:Seabien, G.; Rosinski-Chupin, I.; Mattei, M.G.; Mbikay, M.; Chretien, M.; Seidah, N.G.
Biochemstry 30, 1628-1635, 1991
A:Title: Gene structure and chromosomal localization of plasma kallikrein.
A:Reference number: A39180; MUID:91129236; PMID:1993180
A:Accession: A39180
A:Molecule type: DNA
A:Residues: 1-638 <BEA>
A:Cross-references: UNIPROT:P14272; GB:J05315
A:Note: the authors translated the codon GAG for residue 81 as Gln
R:Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazure
DNA 8, 563-574, 1989
A:Title: The cDNA structure of rat plasma kallikrein.
A:Reference number: A33320; MUID:90091743; PMID:2598771
A:Accession: A33320
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-638 <SEI>
A:Cross-references: GB:M0282; NID:9205010; PIDN:AAA41463.1; PID:9205011
A:Note: part of this sequence, including the amino ends of both the heavy and light chain
R:Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
Biochim. Biophys. Acta 999, 103-110, 1989

```

```

A:Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
A:Reference number: S06851; MUID:90089457; PMID:2597701
A:Accession: S06851.
A:Molecule type: protein
A:Residues: 20-45;391-413 <PAQ>
R:Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazure
DNA Cell Biol. 8, 563-574, 1989
A:Title: The cDNA structure of rat plasma kallikrein.
A:Reference number: I53041
A:Accession: I53041
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-638 <RRS>
A:Cross-references: GB:M58590; NID:9206721; PIDN:AAA42069.1; PID:9206722
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w/
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 115
are linked by one or more disulfide bonds.
C:Genetics:
A:Gene: PK
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inflan
F:1-19/Domain: signal sequence; duplication; status predicted <SIG>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-
F:127,215,308,453,459,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:396/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:433,483,578/Active site: His, Asp, Ser #status predicted

Query Match 10.8%; Score 507.5; DB 1; Length 638;
Best local similarity 26.5%; Pred. No. 1.6e-25;
Matches 183; Conservative 79; Mismatches 220; Indels 209; Gaps 35;

Qy 271 RGSPLVYVNTLSMEBALVOLCGTPSPSYNLTFFHSQVLLITLTNTERHRRPG--E 328
      :|||:
Db 32 RGSGLAIVY---PDAGHC-QKQCTFPRCLFSF-----LAVSPKXETDKRGCKMKE 81
Qy 329 ATFFOLPRMSSCG---GRLRKAQGTENSPYRGHYP-----NIDCTWNIE-----V 372
      :|||:
Db 82 SITGTLPRIRHTGALISGSHLKQCGHQSACHRDYIKGLDMRGSNFNISKNTNIEBQKLC 141
Qy 373 PNNQHVKKRFR---FYLLF-----PGVPACTCPDYVEING--EKYCG----- 411
      :|||:
Db 142 TNNHFCQF--FYVAISAFRPEYRKXCLKLKHSSGTPITSKADNLVSGFSLKSCALSEIG 200
Qy 412 -----ERSQFVNTSNKLTVAFHDSQSTVDGFLAEYLSYSSDPCPGQFCRTGRCI 465
      :|||:
Db 201 CPMDIFQHSADLNVSGVITPDPAVCTICTFPHNCLFTFYTNWETESQRNVCFUKT 229
Qy 466 KRELRCGMADCTHSDLELNCSDAGHQFTCKNFKCKPLF-----WVCDVSV-NDCGDN 517
      :|||:
Db 230 ---VCTFHP--NC-----LFTFYTNWETESQRNVCFUKT 259
Qy 518 SDEQSGSCPA-----QTFRCNGKLSKQCGNGKDCG-----DGS 554
      :|||:
Db 260 TSKSGRPSPIRPOENALISGYSLLTCKRK-----ARPECHFXIVGVAPEGEELNATFVQGA 315
Qy 555 D--EASGPXNVVTCGHTYRCLNGCLSKGNPECKECS-----DGS----- 597
      :|||:
Db 316 DACQETCTK--TIRCOFTYSLLPQDCKAG--C-----KCSLRSLTDSGPTPTTYEAQG 366
Qy 598 -----DEKDCGLRSFTQARVVGTDADGEMPMQVSLHA--LGQGHICGA 643
      :|||:
Db 367 SSGYSRLCKVSVSSDCTTKI-----NARIVGGINSLGEMPMQVSLQVLTGQNMHCGG 421
Qy 644 SLSPNVLVSAHACYIDRGRFRYSPTQWTAFLGLHDQSQ--RSAPGVQERLKRILSHIP 701
      :|||:
Db 422 SIIGRWILTAAHCF--DGIPY--PDVWRITYGILSLSEITWTKTP---FSRIKELIHQ 473

```

Qy 702 FENDFTDYDIALLELEKPAEYSSMWRPCLPDASHVFPAGKAIWVTGHTYOGTGAL 761  
 Db 474 KYKMSSEGSYDIALKLTQPLNTYEFQKPCPSKADNTNTITYNQWVTGWGTTKRGSTQN 533  
 Qy 762 ILQKGEIRVINTQTCENLFPQO-ITPRMVCVGLSGVDSQCGSGGPLSVSEADGRIFG 820  
 Db 534 ILQKATPLPVNEBCQKRDYVITKQMTICAGYKGGIDACKGDSGGL-VCKRSGRWOL 592  
 Qy 821 AGVSWGDCGAGRNKPGVYTRLPFRDMKE 851  
 Db 593 VGITWSGCGARKQPGVYTKVAEYIDWILE 623  
 RESULT 10  
 PLPG  
 plasmin (EC 3.4.21.7) precursor - pig (fragment)  
 N:Alternate names: plasminogen  
 N:Contains: miniplasminogen  
 C:Species: Sus scrofa domestica (domestic pig)  
 C>Date: 07-Sep-1990 #sequence revision 01-Nov-1996 #text\_change 09-Jul-2004  
 C:Accession: S03733; S03737; A25834  
 R:Schaller, J.; Marti, T.; Roesele, S.J.; Kaempfer, U.; Rickli, E.E.  
 R:Schaller, J.; Marti, T.; Roesele, S.J.; Kaempfer, U.; Rickli, E.E.  
 A:Title: Amino acid sequence of the heavy chain of porcine plasmin. Comparison of the ca  
 A:Reference number: S03733  
 A:Accession: S03733  
 A:Molecule type: protein  
 A:Residues: 1-560 <SCH>  
 A:Cross-references: UNIPROT:P06867  
 R:Brunsholz, R.A.; Lersch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg, M.;  
 Eur. J. Biochem. 114, 465-470, 1981  
 A:Title: Comparison of the primary structure of the N-terminal CNBr fragments of human,  
 A:Reference number: S03735; MUID:81212097; PMID:7238497  
 A:Accession: S03737  
 A:Molecule type: protein  
 A:Residues: 1-57 <BRU>  
 R:Marti, T.; Schaller, J.; Rickli, E.E.  
 Eur. J. Biochem. 149, 279-285, 1985  
 A:Title: Determination of the complete amino-acid sequence of porcine miniplasminogen.  
 A:Reference number: A25834; MUID:85203907; PMID:3846533  
 A:Accession: A25834  
 A:Molecule type: protein  
 A:Residues: 450-790 <MAR>  
 C:Function:  
 A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va  
 A:Pathway: fibrinolysis  
 C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology  
 C:Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma; serine prote  
 F:1-790/Product: plasminogen #status predicted <PRO>  
 F:1-77/Domains: plasminogen-related protein precursor homology (fragment) <PLPH>  
 F:1-77/Domains: activation peptide #status predicted <APT>  
 F:78-560/Product: plasmin chain A #status predicted <ACH>  
 F:84-162/Domains: kringle homology <KR1>  
 F:166-243/Domains: kringle homology <KR2>  
 F:256-333/Domains: kringle homology <KR3>  
 F:358-435/Domains: kringle homology <KR4>  
 F:450-790/Product: miniplasminogen #status experimental <MIN>  
 F:461-540/Domains: kringle homology <KR5>  
 F:561-790/Product: plasmin chain B #status experimental <BGH>  
 F:561-783/Domains: trypsin homology <TRY>  
 F:30-54,34-42,84-162,105-145,133-157,166-243,169-297,187-226,215-238,256-333,277-316,305  
 bonds: #status predicted  
 F:602,645,740/Active site: His, Asp, Ser #status predicted  
 Query Match 10.8%; Score 506.5; DB 1; Length 790;  
 Best Local Similarity 23.2%; Pred. No. 2.4e-25;  
 Matches 202; Conservative 100; Mismatches 301; Indels 267; Gaps 36;  
 Qy 106 ENSNSTFVSLASVYKDAL---KLTVSGVPLPGYHHSNATVASEGSVIAYVWSEFS-- 160  
 Db 59 ENSKT---SPLAMRVVPLFEKRIYLSCKTGNGKNYRGRTTSKTKSGVICQKWSVSPH 114

Qy 161 IPOHLVEAEERVAEERVWMLPPRARSLKSFVTSVVAPEPDSKTVORTDNCSFGLHA 220  
 Db 115 IPRYSKEKPLAILEENYCNPNPNDK-----GPMCTTDBET----- 152  
 Qy 221 RGVLEAFETTPGPPSPYPAHACQNALRGDADSVLSLFRSDLASCDERGGDLTVYN 280  
 Db 153 -----RFYCDIPE-----CEDCMHSGEHNYSKSTKWSGIE-----COSWSG----- 192  
 Qy 281 TLSPEPHALVOLCGTYPPSYNLTFHSSQVLLITLITTERA-----HPGEATFOL 334  
 Db 193 -----QSPHAGYLPKSF-PRKULKNYCRN-----PDGEFPPKCTTDPNRMWFCDI 240  
 Qy 335 PRMS-----CGRLRKAQGTN-----SPYYPGHYPNIDCTWN 369  
 Db 241 PRCTTPPTSGPTTGYQCLKKGGENYRGTVASGHTCQWSAQSPKHNRTPENPCK-N 299  
 Qy 370 IEV-----PNNQV-----KYRKFPL-----LFGVPACTP--KDYV 402  
 Db 300 LLENYCRNPDEGTAFCWCTTDSFVRWDYCKIPCGSSTSTHLDAPVPEQTPVAQDCY 359  
 Qy 403 EINGEKYCGERSQFVVTNSNK-ITVRFSDD-----SYDTGFLAYLSYSSDPGPGF 457  
 Db 360 KNGESTYRGTSITITGRKQCSQVSWMTPRHEXTPGNFPNAGLITMYCNPNADKSPWCY 419  
 Qy 458 T-----CRTRGCRKE----- 468  
 Db 420 TTDPRVRWEYCNLKKSETEQVTFNPALAQVPSVEDLSDCMFGNGKRYRGRATTVAG 479  
 Qy 469 LRCDGADCTDHSDELANCSDAGHOFTCKNPKFLFWCDSDVNDGDNDSDEGGSC-- 525  
 Db 480 VPCQEMAAQEPHNSI-FTPETNPRAGLEKNYCR-----NPGDDNGPWCYTT 526  
 Qy 526 -PAQTF-RCNGKCLSKSQCCNGKCGDSDGSDRASPKNVVTCTKHYRCLNGLCISKG 583  
 Db 527 NPQKLFDCVPCVTS-----FDCK-----PKV----- 552  
 Qy 584 NPECDKEDSDSDKDCDCCGLRFTROARVVGTDADGEMPVQVSLHALCGHI CGA 643  
 Db 553 -----EPKCK-----PARVVGCGVSIIPHSWPMQISLRYYRGHFCG 589  
 Qy 644 SLTSPWMLVSAACVYIDDGFRYSPTQWTAFLGLDQORASRPGVDERLRKIIISHPF 703  
 Db 590 TLISPMVLTAKCL-----EKSSPSSYKVIIGAAHE-YHLEGEVQEDIVSLFYEP-- 641  
 Qy 704 NDFTPYDIALLELEKPAEYSSMWRPCLPDASHVFPAGKAIWVTGHTYOGTGALIT 763  
 Db 642 -----SEADIALKLTSSPAVITDKVITPACLPTRYVVAADTACTITMGERTK-CTYGA 696  
 Qy 764 QKGEIRVINTQTCEN-NULLPQITPRMVCVGLSGVDSQCGSGGPLSVSEADGRIFGA 821  
 Db 697 KEARLPEVIEKVCNRYEYLGKVSFMLECAGHLAGIDSCQSGSGPLVCFEKDYIL-Q 755  
 Qy 822 GVSWSGDCGAGRNKPGVYTRLPFRDMKE 851  
 Db 756 GVTSWGLGALPNKPGVYVRSFVTWIEB 785  
 RESULT 11  
 S15468  
 Complement C3b/C4b inactivator (EC 3.4.21.-) precursor - African clawed frog  
 N:Alternate names: C3b/C4b inactivator factor I  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 20-Feb-1995 #sequence revision 20-Feb-1995 #text\_change 09-Jul-2004  
 C:Accession: I51601; S15468  
 R:Kunath-Muglia, L.M.; Chang, G.H.; Sim, R.B.; Day, A.J.; Ezekowitz, R.A.  
 Mol. Immunol. 30, 1249-1256, 1993  
 A:Title: Characterization of Xenopus laevis complement factor I structure--conservation  
 A:Reference number: I51601; MUID:94019415; PMID:7692240  
 A:Accession: I51601  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-613 <KUN>

F;127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimental

F,318-347,340-345/Disulfide bonds: #status predicted  
F,390-391/Cleavage site: Arg-11e (coagulation factor XIIa) #status predicted  
F,434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 10.6%; Score 495; DB 1; Length 638;  
Best Local Similarity 26.8%; Pred. No. 1e-24;  
Matches 183; Conservative 70; Mismatches 209; Indels 222; Gaps 31;

```
QY 328 EATFF-OLPMSSCGGRLRAQGF-----NSPYGHVPPNIDTWNIE-----VPN 374
DB 6 QATFFILFATVSGGCLTQLYENAFPGGVASNTYNAOYCMKCTFHRCILFSLPA 65
QY 375 NQHVKAFKFFYLPEVPAGTCKDY---VEINGEKYCGERSQV-----V 418
DB 66 SSINDMKRFKGFCKDSV-TGTLPKVARTGAVSGHSLKCGHQISACHRDYKGVDRGV 124
QY 419 TSNMNTK-----TYRHFSDQ-----SY 435
DB 125 NFNVSKVSVSECCQKCTNNIRCFPSYATQTFHKAERYNNCLIKYSPGCTPAIKVLSN 184
QY 436 TDGFF-----LAELYSDSDPC-----PGQFCTGRGIRKELRCQGMAD 476
DB 185 VESGFLSKPGALSGHNNIFQHLAFSDVDVARVLTDFAVCRT-----I 230
QY 477 CTDSDELNCSGAGHQFTCKNFKCPLF-----WVCDV-NDGDSNDEG--CSC 525
DB 231 CTYHP--NC-----LFFFTYTNWKIESQRYVCLIKTISEGTPASSST 270
QY 526 PAQFF-----RCSNGKCLSKSQOQCKGKDCGSGDEASCPKNV--TCTK-- 569
DB 271 PQENTISGYLLTCKRLLPEPCHSKIYP--GVDFGSELNVTYKGVNVCQETCTKMRIC 328
QY 570 --HTYRCLNGCLSKGNPECDGKEDCSGSDKDCDGLR-----SFTROA----- 613
DB 329 QEFFYSL-----PEDCK--EKKCKFKLMSMGSTRIAYGTQSSGV 370
QY 614 -----RVGGTDADEGEMQVSLAH--LGQGHICASLISPMVLV 652
DB 371 SLRLCNTGDSVCTTKTSTRIVGCTNSMGEMQVSLQVLTQORHLCGSLIGHQWL 430
QY 653 SAHCVYIDRGFRSDPTQMTAFGLHDQSG--RSAAGVERRLKRIISHPFENDPFEDY 710
DB 431 TAAHCF--DGLPIQD--VWRIYSGLINLSDITQTPSQ--IKELIITHONYKVSNGH 482
QY 711 DIALLELEKPAEYSSMVRPICLPDASHVFPAGKAIWYTGWHTQYGGTGALLQKGEIRV 770
DB 483 DIALIKQALNLTPEPKPICLPBKGDTSTIYTNQWLTGWFGEKEKEINILQKNVPL 542
QY 771 INOTTGENTLPPQ-OITPRMNCVGLSGVDSCQDSCGGLSVSEADGRIFGAGVSKGDG 829
DB 543 VTNEECCKRYQDYKTIQRMVACAGYKEGKDKACKGDSGGL--VCKHNGMRLVGTISWEG 601
QY 830 CAGRNKPGVYTRLLPFRDWIKENT 853
DB 602 CARBEPQVYTKVAEYMDWILEKT 625
```

## RESULT 14

B61545  
plasmin (EC 3.4.21.7) precursor - sheep (fragments)  
N/Alternate names: Plasminogen  
C/Species: miniplasminogen  
C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C/Date: 28-Oct-1994 #sequence\_revision 01-Nov-1996 #text\_change 09-Jul-2004  
C/Accession: B61545; S28200  
R/Schaller, J.; Rickli, E.E.  
Enzyme 40, 63-69, 1988  
A/Title: Structural aspects of the plasminogen of various species.  
A/Reference number: A61545; MUID:8905015; PMID:3168975  
A/Accession: B61545  
A/Molecule type: protein  
A/Residues: 1-37,38-117 <SCH>  
A/Cross-references: UNIPROT:P81286

R/Schaller, J.; Straub, C.; Kaempfer, U.; Rickli, E.E.  
protein Seq. Data Anal. 5, 21-25, 1992  
A/Title: Complete amino acid sequence of ovine miniplasminogen.  
A/Reference number: S28200; MUID:93149995; PMID:1492092  
A/Accession: S28200  
A/Molecule type: protein  
A/Residues: 118-460 <SCH>  
C/Species: Homo sapiens (man)  
C/Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 09-Jul-2004  
C/Accession: A29154; A28434; S66420  
R/Cattell, C.F.; Lyons, A.; Sim, R.B.; Day, A.J.; Harris, T.J.R.  
Biochem. J. 242, 849-856, 1987  
A/Title: Characterization of the primary amino acid sequence of human complement control  
A/Reference number: A29154; MUID:87241401; PMID:2954545  
A/Accession: A29154  
A/Molecule type: mRNA  
A/Residues: 1-583 <CAT>  
A/Cross-references: UNIPROT:P05156; GB:Y00318  
R/Goldberger, G.; Brune, G.A.P.; Rice, M.; Edge, M.D.; Kwiatkowski, D.J.  
J. Biol. Chem. 262, 10065-10071, 1987

```
QY 405 NGEKYCGERSQFV---TSNKKITVRFH-SQSTTDTGFLABYLSYSSDPCQGFIC 459
DB 45 NGQYRGTSSTVTYGRKCKOSWSSMIPHRQKPESEYPNGLTMNYCRNPADKSPWCYT- 103
QY 460 RTGRCIRKE---LRCDGADCTDSDELNCSGAGHQFTCKNFK-----CKPLFWVCD 510
DB 104 -TTPRVRFECNLKAPQAPSVENPPADOMLGIGYGRKATTVAGVPCQD--WAAQE 160
QY 511 VNDGDSNDEGSCPAQTFRCNSGKCLSKSQOQCKGKDCGSDGDEASCPKNVVTCTGH 570
DB 161 PRRHGIPTPE-----TNPRAGLEKRYCRNPD--GD-----VNGPWCYTT 197
QY 571 TYRCLNGCLSKGNPECDGKEDCSGSDSDE-KQDCGLRFTROARVVGSTDADEGEMQ 629
DB 198 NPKRLDYC--DIPCESSFDGKRPKPKC-----PARVAGCVAPHPSPMQ 245
QY 630 VSLHALGQGHICASLISPMVLVSAHCVYIDRGFRYSDPTQMTAFGLHDQSGSAPGV 689
DB 246 VSLRRSRHFCGDTLISEWVLTANCLDSLIG-----PSFTYVLGHAYEAKRA-SV 289
QY 690 QERRRLRIISHPEPNDFTDYDIALLEKPAEYSSMVRPICLPDASHVFPAGKAIWYTG 749
DB 300 QEIPVGRFLPE-----SRADIALKLSSPAVITDEVIACPLSPBNYVADKTVCYTG 353
QY 750 WGHYVGGGALLQKGEIRVINTTCE--NLLPQITPRMNCVGLSGVDSCQDSCG 807
DB 354 WGETQ--GTFGVGLKEARLPVLENKVCNRYEYLNGVKSTELCAQDLAGTDSQDSCG 412
QY 808 PLSSVADGRIFGAGVSMGDCAGRNKPGVYTRLLPFRDWIKE 851
DB 413 PLVCFEKDKTIL-QSVTSMGLGCAKPNKGVYRVSTYPMIEE 455
```

## RESULT 15

A29154  
complement factor I (EC 3.4.21.45) precursor - human  
N/Alternate names: C3b/C4b inactivator  
C/Species: Homo sapiens (man)  
C/Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 09-Jul-2004  
C/Accession: A29154; A28434; S66420  
R/Cattell, C.F.; Lyons, A.; Sim, R.B.; Day, A.J.; Harris, T.J.R.  
Biochem. J. 242, 849-856, 1987  
A/Title: Characterization of the primary amino acid sequence of human complement control  
A/Reference number: A29154; MUID:87241401; PMID:2954545  
A/Accession: A29154  
A/Molecule type: mRNA  
A/Residues: 1-583 <CAT>  
A/Cross-references: UNIPROT:P05156; GB:Y00318  
R/Goldberger, G.; Brune, G.A.P.; Rice, M.; Edge, M.D.; Kwiatkowski, D.J.  
J. Biol. Chem. 262, 10065-10071, 1987







A: Molecule type: protein  
A:Residues: 483-507,'E',509-604 <M13>  
R:Robbins, K.C.; Bernabe, P.; Arzadon, L.; Summari, L.  
J. Biol. Chem. 248, 1631-1633, 1973  
A:Title: The primary structure of human plasminogen. II. The histidine loop of human plasminogen.  
A:Reference number: A92125; PMID:73149248; PMID:4694729  
A:Contents: annotation: active site  
R:Groskopf, W.R.; Summari, L.; Robbins, K.C.  
J. Biol. Chem. 244, 3590-3597, 1969  
A:Title: Studies on the active center of human plasmin. Partial amino acid sequence of a  
A:Reference number: A92048; PMID:69234739; PMID:4240117  
A:Contents: annotation: active site  
R:Trexler, M.; Vail, Z.; Patchy, L.  
J. Biol. Chem. 257, 7401-7406, 1982  
A:Title: Structure of the omega-aminocarboxylic acid-binding sites of human plasminogen.  
A:Reference number: A92382; PMID:82213905; PMID:6919539  
A:Contents: annotation: omega-aminocarboxylic acid binding sites  
R:Vail, Z.; Patchy, L.  
J. Biol. Chem. 259, 13690-13694, 1984  
A:Title: The fibrin-binding site of human plasminogen. Arginines 32 and 34 are essential  
A:Reference number: A92458; PMID:85054794; PMID:6094526  
A:Contents: annotation: fibrin binding site; omega-aminocarboxylic acid binding site  
J:Gao, Y.; Ji, R.W.; Davidson, D.; Schaller, J.; Marti, D.; Soehndel, S.; McCance, S.G.;  
J. Biol. Chem. 271, 29461-29467, 1996  
A:Title: Kringel domains of human angiotensin. Characterization of the anti-proliferative  
A:Reference number: A58811; PMID:97067211; PMID:8910613  
A:Contents: annotation  
R:Iljnen, H.R.; Uggul, F.; Bahl, A.; Collen, D.  
Biochemistry 37, 4699-4702, 1998  
A:Title: Generation of an angiotensin-like fragment from plasminogen by streptolysin-1 (M  
A:Reference number: A58812; PMID:9548733; PMID:9548733  
A:Contents: annotation  
R:Iljnen, H.R.; Uggul, F.; Bahl, A.; Collen, D.  
Biochemistry 37, 4699-4702, 1998  
A:Title: Generation of an angiotensin-like fragment from plasminogen by streptolysin-1 (M  
A:Reference number: A51341; PMID:1181184  
A:Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 376-454  
R:Tulinsky, A.; Wu, T.P.  
submitted to the Brookhaven Protein Data Bank, July 1991  
A:Reference number: A51488; PMID:125284  
A:Contents: annotation: X-ray crystallography, 2.25 angstroms, residues 375-454  
R:Wu, T.P.; Tulinsky, A.  
submitted to the Brookhaven Protein Data Bank, August 1993  
A:Reference number: A51911; PMID:1181184  
A:Contents: annotation: X-ray crystallography, 2.48 angstroms, residues 102-181  
R:Padmanabhan, K.; Tulinsky, A.  
submitted to the Brookhaven Protein Data Bank, April 1994  
A:Reference number: A52408; PMID:1181184  
A:Contents: annotation: X-ray crystallography, 2.25 angstroms, residues 377-454  
R:Tulinsky, A.; Mathews, I.I.  
submitted to the Brookhaven Protein Data Bank, December 1995  
A:Reference number: A65244; PMID:1181184  
A:Contents: annotation: X-ray crystallography, 2.1 angstroms, residues 102-181  
R:Mulichak, A.M.; Tulinsky, A.; Ravichandran, K.G.  
Biochemistry 30, 10576-10588, 1991  
A:Title: Crystal and molecular structure of human plasminogen kringel 4 refined at 1.9 Å  
A:Reference number: A58819; PMID:92031502; PMID:1657148  
A:Contents: annotation  
R:Wu, T.P.; Padmanabhan, K.; Tulinsky, A.; Mulichak, A.M.  
Biochemistry 30, 10589-10594, 1991  
A:Title: The refined structure of the epsilon-aminocarboxylic acid complex of human plasminogen  
A:Reference number: A58818; PMID:92031503; PMID:1657148  
A:Contents: annotation  
R:Wu, T.P.; Padmanabhan, K.; Tulinsky, A.; Westbrook, M.  
Biochemistry 31, 270-279, 1992  
A:Title: Crystal structure of the kringel 2 domain of tissue plasminogen activator at 2.4 Å  
A:Reference number: A39483; PMID:92118803; PMID:1310033  
A:Contents: annotation: X-ray crystallography, 2.4 angstroms  
R:Stew, B.; Teeter, M.M.; Whitlow, M.; Yamano, A.  
submitted to the Brookhaven Protein Data Bank, June 1995

	A:Reference number: A65980; PDB:1KRN
	A:Contents: annotation; X-ray crystallography, 1.67 angstroms, residues 376-454
	R:Rejante, M.; Llinas, M.
	submitted to the Brookhaven Protein Data Bank, August 1996
	A:Reference number: A65803; PDB:1HPJ
	A:Contents: annotation; conformation by (1)H-NMR, residues 103-181
	R:Rejante, M.; Llinas, M.
	submitted to the Brookhaven Protein Data Bank, August 1996
	A:Reference number: A65804; PDB:1HPK
	A:Contents: annotation; conformation by (1)H-NMR, residues 103-181
	R:Rejante, M.R.; Llinas, M.
	Eur. J. Biochem. 221, 927-937, 1994
	A:Title: (1)H-NMR assignments and secondary structure of human plasminogen kringles 1.
	A:Reference number: A63645; PMID:94237157; PMID:8181475
	A:Contents: annotation; conformation by (1)H-NMR, residues 96-184
	R:Rejante, M.R.; Llinas, M.
	Eur. J. Biochem. 221, 939-949, 1994
	A:Title: Solution structure of the epsilon-aminohexanoic acid complex of human plasminogen
	A:Reference number: A58817; PMID:94237158; PMID:8181476
	A:Contents: annotation; conformation by (1)H-NMR
	C:Comment: Plasminogen is synthesized by the kidney and is present in plasma and many other
	C:Comment: Plasminogen is converted to plasmin by plasminogen activators (see PIR:UKHU at
	d PIR:FGUGB).
	C:Comment: C:Comment: Plasmin is inactivated by alpha-2-antiplasmin (see PIR:ITHU2) immediately after
	its release from the fibrinogen-lysine complex. Plasminogen is converted to plasmin by plasminogen
	activators (see PIR:UKHU at d PIR:FGUGB).
	C:Comment: Streptolysin 1 (see PIR:KCHU1) acts on plasminogen to produce angiotensin. To
	day, solid tumors.
	C:Genetics:
	A:Gene: GDB:PLG
	A:Cross-references: GDB:119498; OMIM:173350
	A:Map position: 6q26-6q27
	A:Introns: 17/1, 62/2, 98/1, 136/2, 183/1, 223/2, 263/1, 317/2, 366/1, 419/2, 480/1, 529/
	2
	C:Function:
	A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a variety
	of tissues; the walls of the graafian follicle; also activates the urokinase-type plasminogen acti
	A:Pathway: fibrinolysis
	C:Superfamily: plasmin; kringles; plasminogen-related protein precursor homology <PLPH>
	C:Keywords: angiotensin inhibitor; blood; duplication; fibrinolysis; glycoprotein; hydro
	F:1-96/Domain: plasminogen-related protein precursor homology <PLPH>
	F:1-19/Domain: signal sequence #status predicted <SIG>
	F:20-810/Product: plasminogen #status experimental <PRO>
	F:20-96/Domain: activation peptide #status experimental <APT>
	F:97-466/Product: angiotensin #status experimental <AST>
	F:97-580, 581-810/Product: plasmin #status experimental <MAT>
	F:97-580/Domain: plasmin chain A #status experimental <CHA>
	F:103-181/Domain: kringles homology <KR1>
	F:185-262/Domain: kringles homology <KR2>
	F:275-352/Domain: kringles homology <KR3>
	F:377-454/Domain: kringles homology <KR4>
	F:481-550/Domain: kringles homology <KR5>
	F:550-580, 581-810/Product: microplasmin #status experimental <MPT>
Query Match	10.4%; Score 488; DB 1; Length 810;
Best Local Similarity	24.5%; Pred. No. 4e-24;
Matches 223; Conservative 103; Mismatches 296; Indels 288; Gaps 49;	
QY	102 VDAYENSNSTEEVSLASKYKDAKLILYSGVPLGPHKBEAYAFSGSVIAYWSEF-- 159
DB	22 LDDVYNTGASLFSVYKQ-----LGAGSTECACKEED-----EEFTC 61
QY	160 -SIPQHLVEARVWAEEVWMLPPRARSILKSFVSVVAFPYDSK-----TYQRT 209
DB	62 RAYQHSSEQQGVVMAENKSSIIITRMV--VLEPKYVLSSEKTKGNKYNRTSKT 118
QY	210 QDNQSCSFGILHARGVEIMRFTT-----PGFPDSVPYAH-----RCQ-----WALRGD 251
DB	119 KN-----GIRCOKMSSTPHRPRSPATHPEGLSEENYCRNPDNDPGGWCYTTD 168
QY	252 ADSVSLTPFRSDLASCDER-----GSDL-VIVYNTLSPE-----PHALVOLGQV 297
DB	169 PER---RDYDILLCEECCHNGSCGENYDGKISKMSGLTECAMDSQSPA---HGYT 220









F:481-560/Domains: kringles homology <KR5>  
 F:581-803/Domains: trypsin homology <TRY>  
 F:49-73, 53-61, 103-181, 124-164, 152-176, 185-262, 168-316, 206-245, 234-257, 275-352, 296-335, 32  
 Bonds: #status predicted  
 F:622,665,760/Active site: His, Asp, Ser #status predicted

Query Match 10.2%; Score 477; DB 2; Length 810;  
 Best Local Similarity 24.5%; Pred. No. 2,1e-23;  
 Matches 226; Conservative 96; Mismatches 286; Indels 312; Gaps 50;

QY 102 VDVAENSNSTFVSLASKVKALKLYSGVPLPGYHKEASAVTAFSEGSVIAVYSEF- 159  
 DB 22 LDDVNTTGAISLFIITKKQAGSI-----ECCAACEE-----EEFTC 61  
 QY 160 -SIQHLVEAEERVAERVMVLPRAASLKSFVTVSVAP----- 200  
 DB 62 RSFOYHSKEQOCVMAENR-----KSLVFNRDVLVEKKVYLSECKTGNGKNYRG 113  
 QY 201 TDSKTVOITQDNCSFGIHAARVELMRETT-----PCPPDSPYPAHA-----RCQ----- 245  
 DB 114 TMSKI--RT-----GITCQKWSSTSPHRTSPATHSSEGLEENYCNPNPDG 159  
 QY 246 ---WALRGDADSVLSLTFRSPDLASCDER-----GSDL-VTVNTLSPEM-----PH 288  
 DB 160 QGPMCYTTDPE-----REDYCDIPECEDECHMGSGENYDGKISKTMGLFCQAWDSQSPH 215  
 QY 289 ALVOLCGTYR---PSYNTLTFHSSQNVLLITLITRER-----HGEAFATFQLPMS- 338  
 DB 216 A---HGATPEKFNKILKKNYCN-----PGDEPPMCFTTDPNKMELCDI PRCTT 264  
 QY 339 ---SCGRLRAQGTFSNPPYGHYPNID--C-TWNIENVNNOHVAREKFFYLEPG 391  
 DB 265 PPSSGPTYOCLKKTGEN--YRGDAVAVTSGHTCHGSAQCPHTHN----- 308  
 QY 392 VPACTCPDYVEIN-GEKYC---GERSQFVTVSNKITYR----- 428  
 DB 309 ---RTPENFPCNKLDEYCNPNPGEKAPCYTINSQ---VRMEYCKIPSCSSPVSTEP 361  
 QY 429 -----FHSD-QSTYDNGFLAEVLSYDSDPCPOQFCTRCRCKEIRC 471  
 DB 362 LDPTAPPELFPVVECHYGDQSYRG-----SSTTTGK-----KC 398  
 QY 472 DGMADCTDHSDELNCSCDAGHQFTCKNFKCPFLFVVCDSVNDGDNDSQSC---PA 527  
 DB 399 QSNWSMTPH-----WHEKTPENPNAGL-----TMYCNPNADKCPKCFETDPS 443  
 QY 528 QTFPCSNKCLSKSQCGNG-----KDCG--GPGSDASCPKYNV 564  
 DB 444 VRMEYCNULK-----KCSGTGSAVAAPPAQLPDAETPSBEDCMFGNGKGYRG-KKATY 496  
 QY 565 VTCT-----KHTYRCLN-----GL-----CLSK-----GNPEC--DKEKDCSDSD 598  
 DB 497 VTGPRCOEMAQEPHSHRIFTPTBNPRAGLEKNYCRNPDGDPVGGPMCTTNPRLKLFYCD 556  
 QY 599 EKDC-----DCGLRSFTRO---ARVGGTDADEGEPMQVSLHAGQHLGASLSPNW 650  
 DB 557 VPQDASSPDCGQKQVPRKCPGRVAVGCAVYPSWPKQISLRLRLGKHFCCGLISBEM 616  
 QY 651 LVSAAHCYIDRGRYSPTQMTAFGLHDSQSRAPQVORRLKRIISHPFENDFTDY 710  
 DB 617 VLTAAHCL-----EKSSRPSFYKYLGAH-REVLHPRVQIEIVSKMSEP-----ARA 664  
 QY 711 DIALLEKPAEYSMRPTCLPDASHVFPAGKAIWVTGWHGTQYGGGAILLOKGEIRV 770  
 DB 665 DIALKISSPAIITDKVIAPLCPSPNYVADRTCEFTIGMETG-GTYGAGLLEARLPIV 723  
 QY 771 INQTTCE--NLTPQOITPRMVCVFLSGVDSGQSDSGPISYEADGRIRGAGVNSWD 828  
 DB 724 IENKVCNRYEFLNLTGVTTELCAHLAGTDSGQSDSGPLVCFBKXKYL--QGVTSWGL 782  
 QY 829 GCAGRNKPGVYTRLLPFRDWIK 850  
 DB 783 GCARNKPGVYTRVRSFVTWIE 804

## RESULT 24

A59271  
 Ra-reactive factor (EC 3.4.21.-) 2 precursor - human  
 N:Alternate names: mannose binding protein-associated serine proteinase 2 (MASP-2)  
 C:Species: Homo sapiens (man)  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
 C:Accession: A59271  
 R:Title: S.; Vorup-Jensen, T.; Stover, C.M.; Schwaible, W.J.; Laureen, S.B.; Poulsen, K.  
 Nature 386, 506-510, 1997  
 A:Title: A second serine protease associated with mannan-binding lectin that activates C  
 A:Reference number: A59271, MUID:97242412, PMID:9087411  
 A:Accession: A59271  
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-686 <JEN>  
 A:Cross-references: UNIPROT:O00187; GB:Y09926; NID:G4007626; PIDN:CAA71059.1; PID:G40076;  
 A:Experimental source: tissue liver  
 A:Note: submitted to GenBank, December 1996  
 A:Note: parts of this sequence, including the amino end of the mature protein, were deter  
 C:Genetics:  
 A:Gene: GDB:MASP2  
 A:Cross-references: GDB:6071500  
 A:Map position: 1p36.2-1p36.3  
 C:Superfamily: complement-activating serine proteases C1r/C1s/MASP; C1r/C1s repeat homol  
 C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; hydrolase; serine pr  
 F:1-15/Domains: signal sequence #status predicted <SIG>  
 F:16-44,445-686/Product: Ra-reactive factor 2 #status predicted <MAT>  
 F:19-134/Domains: C1r/C1s repeat homology <C1R1>  
 F:142-180/Domains: EGF homology <EGF>  
 F:184-293/Domains: C1r/C1s repeat homology <C1R2>  
 F:300-361/Domains: complement factor H repeat homology <FH1>  
 F:366-430/Domains: complement factor H repeat homology <FH2>  
 F:445-679/Domains: trypsin homology <TRY>  
 F:72-90,142-156,152-165,167-180,184-211,241-259,300-348,328-361,366-412,396-430,434-552,  
 F:158/Modified site: erythro-beta-hydroxyasparagine (asn) #status predicted  
 F:444-445/Cleavage site: Arg-11e (autolytic) #status predicted  
 F:483,532,633/Active site: His, Asp, Ser #status predicted

Query Match 10.1%; Score 475; DB 1; Length 686;  
 Best Local Similarity 25.4%; Pred. No. 2.3e-23;  
 Matches 185; Conservative 88; Mismatches 275; Indels 180; Gaps 30;  
 QY 227 RFTTPGPPDPYPAHARCQALRGDADSVLSLTFRSPDLASCBERSDLYTVNTLSPE 266  
 DB 29 RLASPGFP-GEYANDQERRWTLTAPPGYRLVLTFTHEDELSHLCEYDFVLSGAK-- 84  
 QY 287 PHALVOLCG-----TYPSPYNTLTFHSSQNVLLITLITRERHP--GFEATFQLPMS 339  
 DB 85 --VLATTCGGEISTDERAPKQDFFYLSGSLDITFRSDYSENEPFTGFEA-FYAAEDIDE 141  
 QY 340 C-----GGLRAQGTFSNPPY 356  
 DB 142 QVAPGEAPTCDHCHNLGFGYCSGAGVVLHNRKRTCSALCSGVFTQSGELSSPY 201  
 QY 357 PGHYPNIDCTWNIENVNNOHVAREKFFYLEPVPAGCPDYVEINEK-----YCG 411  
 DB 202 PRPYPKLSCTYSLSEGEFVILDFESPDVTH--PELCPVDPLKIQDRHEHGF 260  
 QY 412 ERSQFVTVSNKITYAFHSDQSYTDGFLAEVLSYDSDPCP-----GQFTCRGTGCR 466  
 DB 261 KTLPHRLTETSNVTYITFVDESDHTGKINTSTAHACPYMAPRNGVSVQAKYL 320  
 QY 467 KELRCDGMACTDHSDELNCSCDAGHQFTCKNFKCPFLFVVCDSVNDGDNSEQ---GC 523  
 DB 321 K-----DSFISFCETGYELGQHLPLKSFYAVCK-----DGSWDRMPAC 361  
 QY 524 SCRAQTRGCSNGKCLSKSQCGNGKDCGSDASCPKYNVYCTGHT-----YRCLN 576  
 DB 362 SIY-----DCGP--PDLPSGRVYITIGPGVTVYKAVIYQSCSE 398  
 QY 577 GLCISKNGPECDKEDC-SDG-----SDEKD---CD--CGLRSFTROARVVGSTDADEGR 625







A>Title: Genomic amplification with transcript sequencing.  
 A:Reference number: 159529; MUID:88127096; PMID:3340835  
 A:Accession: 159529  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 230-359 <RE2>  
 A:Cross-references: GB:M19063; NID:g182622; PIDN:AAA52456.1; PID:g182623  
 R:Agarwala, K.L.; Kawabata, S.; Takao, T.; Murata, H.; Shimonishi, Y.; Nishimura, H.; Iw  
 Biochemistry 33, 5167-5171, 1994  
 A>Title: Activation peptide of human factor IX has oligosaccharides O-glycosidically lin  
 A:Reference number: A54255; MUID:94227047; PMID:8172892  
 A:Accession: A54255  
 A:Molecule type: protein  
 A:Residues: 'D', 204, 'X', 206-211, 212, 'D', 214, 'X', 216-221, 'D', <AGA>  
 A>Note: the residues designated 'X' were determined to be threonine bound to carbohydrate  
 R:Di Sciopio, R.G.; Kurachi, K.; Davie, E.W.  
 J. Clin. Invest. 61, 1528-1538, 1978  
 A>Title: Activation of human factor IX (Christmas factor).  
 A:Reference number: A18483; MUID:78194509; PMID:659613  
 A:Contents: annotation; activation; active site; carbohydrate binding  
 R:McGraw, R.A.; Davie, E.W.; Noyes, C.M.; Graham, J.B.; Roberts, H.R.; Stafford, D.W.  
 Am. Soc. Hematol. Abstr. 64 (Suppl. 1), 262a, 1984  
 A:Reference number: A37569  
 A:Contents: annotation  
 A>Note: 194-Thr was also found  
 R:Mortle, T.; Isaacs, B.S.; Esmon, C.T.; Johnson, A.E.  
 J. Biol. Chem. 259, 5698-5704, 1984  
 A>Title: Derivatives of blood coagulation factor IX contain a high affinity Ca2+-binding  
 A:Reference number: A37543; MUID:84185715; PMID:6425296  
 A:Contents: annotation; calcium binding  
 R:Mortle, T.; Isaacs, B.S.; Esmon, C.T.; Johnson, A.E.  
 J. Biol. Chem. 260, 2583-1985  
 A:Reference number: A37544  
 A:Contents: annotation; calcium binding, correction  
 R:Bentley, A.K.; Reed, D.J.G.; Rizza, C.; Brownlee, G.G.  
 Cell 45, 343-348, 1986  
 A>Title: Defective propeptide processing of blood clotting factor IX caused by mutation  
 A:Reference number: A37545; MUID:86189947; PMID:3009023  
 A:Contents: annotation; signal sequence cleavage site  
 R:Stehiro, K.; Kawabata, S.I.; Miyata, T.; Takeya, H.; Takamatsu, J.; Ogata, K.; Kamiya, J.  
 J. Biol. Chem. 264, 21257-21265, 1989  
 A>Title: Blood clotting factor IX B(M) Nagoya: substitution of arginine 180 by tryptophan  
 A:Reference number: A30622; MUID:90078229; PMID:2592373  
 A:Contents: annotation; sequence of mutant B(M) Nagoya  
 A>Note: carboxylation, glycosylation, and cleavage sites  
 R:Baron, M.; Norman, D.G.; Harvey, T.S.; Hanford, P.A.; Mayhew, M.; Tse, A.G.D.; Brownle  
 submitted to the Brookhaven Protein Data Bank, November 1991  
 A:Reference number: A51252; PDB:1IXA  
 A:Contents: annotation; conformation by (1)H-NMR, residues 92-130  
 A>Note: recombinant form expressed in yeast  
 C:Comment: Factor IX is activated by factor XIa, which excises the activation peptide pr  
 C:Comment: The gamma-carboxyglutamic acid residues arise by posttranslational, vitamin K  
 C:Comment: Calcium binds to the gamma-carboxyglutamic acid (Gla) residues and, with strc  
 C:Genetic:  
 A:Gene: GDB:F9  
 A:Cross-references: GDB:119900; OMIM:306900  
 A:Map position: Xq27.1-Xq27.2  
 A:Functions: 30/1; 84/2; 93/1; 131/1; 174/1; 241/3; 280/1  
 C:Function:  
 A:Description: catalyzes the proteolytic activation of coagulation factor X in the prese  
 A:Pathway: blood coagulation intrinsic pathway  
 C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology  
 C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglutan  
 F:1-28/Domain: signal sequence #status experimental <SIG>  
 F:29-46/Domain: propeptide #status experimental <PPT>  
 F:31-91/Domain: Gla domain homology <Gla>  
 F:47-119/Product: coagulation factor IXa light chain #status experimental <ALC>  
 F:97-119/Domain: EGF homology <EGF>  
 F:134-170/Domain: EGF homology <EG2>  
 F:192-226/Domain: activation peptide #status experimental <ACT>  
 F:227-461/Product: coagulation factor IXa heavy chain #status experimental <AHC>  
 F:227-454/Domain: trypsin homology <TRY>  
 F:53,54,61,63,66,67,72,73,76,79,82,86/Modified site: gamma-carboxyglutamic acid (Glu) #

F:64-69,97-108,102-117,119-128,134-145,141-155,157-170,178-335,252-268,382-396,407-435/D  
 F:99/Binding site: carbohydrate (Ser) (covalent) #status experimental  
 F:110/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental  
 F:151-192/Cleavage site: Arg-Ala (coagulation factor Xla) #status experimental  
 F:203,213/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:205,215/Binding site: carbohydrate (Thr) (covalent) #status experimental  
 F:226-227/Cleavage site: Arg-Val (coagulation factor Xla) #status experimental

Query Match 10.1%; Score 474.5; DB 1; Length 461;  
 Best Local Similarity 28.5%; Pred. No. 1.0e-23;  
 Matches 139; Conservative 66; Mismatches 169; Indels 113; Gaps 19;

399 KDVEINGEKYC-GERSQFVNTSNKITYRFRHSDSYTPGFLAEIYSDSDPCPGQF 457  
 Db EEVYQGLERECHEKCSFEAREVENTER-----TTERWKQYVSDQCESNP--- 101  
 QY 458 TCRGRCIRKELKCDGWADCTHSDBLNCSQAGHQ-----FTC-KKKCKPLFWVC 508  
 Db 102 -----CLNGSCKDIDINSYECWCPGFEGKNCLELDTVCNKRCEQF---- 144  
 QY 509 DSVNDCGDNDEQ-GGSC-----PAQFRGNGKCLSKSQCGNCKDCCGDG 553  
 Db 145 -----CKNSADKNVVCSTEGYRLAENQKSCBAVPPCGRVSVQTSKYTRA----- 192  
 QY 554 SDPASCPRVAVVCTKRTYRCLNGLCLSKGNPCDCKEDSDSDCKDCDGLRSPTRQA 613  
 Db 193 --EAVPPDVYVNSTE-----AETLDNITO-----SQSFNDFT 225  
 QY 614 RVVGGTDADGEMPMQVSLHALQGHICGASLISPMWLSAHCYIDRGFRYSDDPTQWT 673  
 Db 226 RVVGGEDAKGQGPWQVVLNGKYDA--FCGGSIVNEKIVVAHCV--ETGVK-----T 276  
 QY 674 AFGLDHQSRSAPGVQERLKLRIHPFNDR--TFDVIALLELEKPMESVMRPLC 721  
 Db 277 VVAGEH-IEETHTGKRVIRIIPHNNAALINKNHIDALEDEPLVLSVTPIC 335  
 QY 732 LPDASHV-----PPAKAIWVTGMHTQVGTGALLQKKEIRIVNOTTCENLLPOQITP 786  
 Db 336 IADKEYNITLTKGSG---YSGMGVRFHKGRLAVQYRLVLDVRLATLRSTKFTIVN 392  
 QY 787 RMCVGFSGVDSQCGDSGSPISVADGRITGAGVSMGDCACGRNKGVTYRLPLFR 846  
 Db 393 NMCAGHFGHGRSDSCQDSGSP-HVTEVESTFLTGIIISWEGECAMKKGKGIYKVSRYV 451  
 QY 847 DWIKENT 853  
 Db 452 NWIKENT 458

RESULT 26  
 KFHUI  
 coagulation factor Xla (EC 3.4.21.27) precursor [validated] - human  
 N:Alternate names: antithrombophilic factor C; plasma chromoplasmin antecedent  
 C:Species: Homo sapiens (man)  
 C:Date: 13-Aug-1986 #sequence revision 26-May-1994 #text\_change 09-Jul-2004  
 C:Accession: A27431; A00920; A37940  
 R:Aakari, R.; Davie, E.W.; Chung, D.W.  
 Biochemistry 26, 7221-7228, 1987  
 A>Title: Organization of the gene for human factor XI.  
 A:Reference number: A27431; MUID:88107663; PMID:2827746  
 A:Accession: A27431  
 A:Molecule type: DNA  
 A:Residues: 1-625 <ASA>  
 A:Cross-references: UNIPROT:P03951; GB:M18295  
 A>Note: the sequence shown follows the authors' translation  
 R:Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.  
 Biochemistry 25, 2417-2424, 1986  
 A>Title: Amino acid sequence of human factor XI, a blood coagulation factor with four tar  
 A:Reference number: A00920; MUID:86243360; PMID:3636155  
 A:Accession: A00920  
 A:Molecule type: mRNA  
 A:Residues: 1-625 <PU>  
 A:Cross-references: GB:M13142; NID:g182832; PIDN:AAA52487.1; PID:g182833

R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.  
Biochemistry 30, 2056-2060, 1991  
A;Title: Location of the disulfide bonds in human coagulation factor XI: the presence of  
A;Reference number: A37940; MUID:9115017; PMID:1998667  
A;Accession: A37940  
A;Molecule type: protein  
A;Residues: 28-33/35-49, 'X', 51-55, 'X', 57-63/70-75, 'X', 77-79/107-109, 'X', 111-112/132-139,  
'X', 280-282, 'X', 284/285-297/313-316, 'X', 318-319/320-326, 'X', 328-330, 'X', 347-349/373, 'X', 379  
C;Comment: The proenzyme consists of two identical chains linked by one or more disulfide  
he active site, and a heavy chain, which associates with high molecular weight (HMW) kin  
C;Genetics:  
A;Gene: GDB:Fl1  
A;Cross-references: GDB:119891; OMIM:264900  
A;Map position: 4q35-4q35  
A;Intons: 19/1; 73/2; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1; 52  
C;Function:  
A;Description: catalyzes the proteolytic activation of coagulation factor IX  
A;Pathway: blood coagulation intrinsic pathway  
C;Superfamily: coagulation factor XI; trypsin homology  
C;Keywords: blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hydric  
F;1-18/Domain: signal sequence #status predicted <SIG>  
F;19-387/Product: coagulation factor Xla heavy chain #status experimental <HCH>  
F;19-108/Domain: apple repeat <AP1>  
F;109-198/Domain: apple repeat <AP2>  
F;199-288/Domain: apple repeat <AP3>  
F;290-379/Domain: apple repeat <AP4>  
F;388-625/Product: coagulation factor Xla light chain #status experimental <LCH>  
F;388-618/Domain: trypsin homology <TRY>  
F;20-103/514-581, 571-599/Disulfide bonds: #status predicted  
F;28/Disulfide bonds: interchain #status experimental  
F;46-76, 50-56, 110-139, 136-165, 140-146, 200-283, 226-255, 230-236, 291-374, 317-346, 321-327, 38  
F;90, 166, 353, 450/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;339/Disulfide bonds: interchain #status predicted  
F;387-388/Cleavage site: Arg-116 (coagulation factor XIla) #status experimental  
F;431, 480, 575/Active site: His, Asp, Ser #status predicted  
F;491/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 10.1%; Score 471; DB 1; Length 625;  
Best Local Similarity 35.4%; Pred. No. 3.8e-23;  
Matches 118; Conservative 42; Mismatches 117; Indels 56; Gaps 13;

QY 554 SDEASCEKV--NVTCTKHYRLCLNGCLSKNCECKGKDC-----SDSDEK----- 600  
Db 313 SHBA-CQKLTNAKRCFFTYPAQASC-----NGKKCKYKLSNNSPTTILHGRG 364  
QY 601 -----DCCGLRSTROARVVGTTADDEGEMPOVSHALG--QGHICGASLI 646  
Db 365 GISGYTLRLCKMNECTTKI---KPRIVGSTASVSGMWPQVTLHTSPQRHLGCGSII 421  
QY 647 SPNWLVAHACYIDDRGFYSPTPTWATFLGLHDQO---RSAPGVQERRLRKRIISHP 702  
Db 422 GNMWILTAHCF-----YGVESPRLILRVSGILNQSIKEDTSFPVQO-----IIHDQ 471  
QY 703 FNDFTPYDIALLELEKPAEYSSWVRICPDASHVPACKAIWYTWGHTQYGGTAL 762  
Db 472 YKMAESGYDIALLELTFTVYVTSQRIICPSKGRDNVYITDQWVTCWGRKADKIQNT 531  
QY 763 LQGEIRVINQTTCE--NLTPQQTTPRMVCVFLSGVDVSCQSGGGLSVEADGRIFG- 820  
Db 532 LQAKAKPIVAVNECQKRYRHKHITKIKICAGYREBGDACKGDSGGFLSC--KENEVWHL 569  
QY 821 AGVAVSGDAGAGNKKPGVYTRLPFRDIKENT 853  
Db 590 VGITSMGECAGQRRERPGVTVNVEYVDILKKT 622

RESULT 27  
S40007  
trypsin (BC 3.4.21.4) precursor - African malaria mosquito  
C;Species: Anopheles gambiae (African malaria mosquito)  
C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
C;Accession: S40007  
R;Muller, H.

submitted to the EMBL Data Library, June 1993  
A;Reference number: S40003  
A;Accession: S40007  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-275 <MOE>  
A;Cross-references: UNIPROT:P35037; EMBL:Z22930; NID:g410323; PIDN:CAA80517.1; PID:g41032;  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; protein digestion; serine proteinase  
F;49-269/Domain: trypsin homology <TRY>  
F;89, 114, 230/Active site: His, Asp, Ser #status predicted

Query Match 10.0%; Score 469; DB 2; Length 275;  
Best Local Similarity 39.9%; Pred. No. 2e-23;  
Matches 97; Conservative 39; Mismatches 89; Indels 18; Gaps 6;

QY 614 RVVCGTDADGEPMQVSHALGQGHICGASLISPMWLVSAHACYIDDRGFRSDPTQWT 673  
Db 48 RIVGGFEIDVSETPYVVSQYF--NSHRCGQSVLNSKMLTAACTVN-----LQPSILA 100  
QY 674 APLGLHDQSORAPGVQERRLRKRIISHPFNDFTPYDIALLELEKPAEYSSWVRICLP 733  
Db 101 VRLG-----SSRLASGCTVVRVARVLEHRYVDDSTIDYDSIMLETFELTFSDVVGVSIP 156  
QY 734 DASHVPACKAIWYTWGHTQYGGTALILQGEIRVINQTTCENLIPQO--ITPMPCV 791  
Db 157 EQDEAVEDGTMTVTVSQWGVQSAEENALIRANIPVNAQECTTIAYSSGGITDMLCA 216  
QY 792 GFLSGGVDSQCGDGGSLSVENDGRIFGAGVSWGDGAGRNKPGVYTRLPFRDIK 851  
Db 217 GYRSGKDKACQDGGSL--VVDGKL--VGAVSWGFCAMPQYPGVYTRVAVRWVRE 271  
QY 852 NTG 854  
Db 272 NSG 274

RESULT 28  
146260  
plasmin (BC 3.4.21.7) precursor - western European hedgehog  
C;Species: Erinaceus europaeus (western European hedgehog)  
C;Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 09-Jul-2004  
C;Accession: I46260  
J;Lawn, R.M.; Boonmark, N.W.; Schwartz, K.; Lindahl, G.E.; Wade, D.P.; Byrne, C.D.; Fong,  
J. Biol. Chem. 270, 24004-24009, 1995  
A;Title: The recurring evolution of Lp(a): Insights from cloning of hedgehog apolipoprote  
A;Reference number: I46259; MUID:96025778; PMID:7592597  
A;Accession: I46260  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-810 <LAW>  
A;Cross-references: UNIPROT:Q29485; EMBL:U33171; NID:g1046360; PID:g1046361  
C;Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology;  
C;Keywords: hydrolase; serine proteinase  
F;1-96/Domain: plasminogen-related protein precursor homology <PLPH>  
F;103-181/Domain: kringle homology <KR1>  
F;185-262/Domain: kringle homology <KR2>  
F;275-352/Domain: kringle homology <KR3>  
F;379-456/Domain: kringle homology <KR4>  
F;482-561/Domain: kringle homology <KR5>  
F;582-803/Domain: trypsin homology <TRY>

Query Match 10.0%; Score 468; DB 2; Length 810;  
Best Local Similarity 28.1%; Pred. No. 8e-23;  
Matches 158; Conservative 62; Mismatches 192; Indels 150; Gaps 27;

QY 405 NGEKTCGERSQGVYVTSNSKTIYFRHSDS-----YTDTGFLAFLSYDSDPQP 454  
Db 279 NGEHYQGN-----VAVTVGLTCQRFWGEQSPRHRTPTENYPCKNLDENYCRNPDEPAP 333  
QY 455 GQPTCTGTCIRKEL-----KCDGMADCTPHSDS-----LNGSCDAGHOF----- 494  
Db 334 WCFT--TNSVWRPECKIPDCVSSASBETHSDAPVIVPEQOTPVVQECTQAGNQTIRGTS 391

QY 495 --TCNKCKPLFV-----CD-SVNDGDNDEQSC-----PAQTR 531  
Db 392 STTTGKKCP--WTSRPHRSHKTPENYPADLTMMYCRNPDGDKMCTTDPVSWE 449  
QY 532 -CSNGKCL-----SKSQOC---NGKDDCGDSDEAS----- 558  
Db 450 FCNKKCSGTEMATNNSPVGVSSASBDDCLIDNGKVRGKATGAGTPCGAMAAQ 509  
QY 559 -----CPKVVVCTKHTYRC-----LNGICLSKNPE-----CDGKEDCSGSD 599  
Db 510 EPHRHSIFTEBETNRADLOENY--CRNPDGANGPCYTTNPRKLFYCDIPHCYSPSS-- 566  
QY 600 KDDCCGARSFTQARV-----VGGTDADGEMPMOVSLHALGOGHICGASLISPNWL 651  
Db 567 -ADCG-----KRVKPKCCPRGVGCVAHPSMPVQSLRRPQ--HFCGGLISPMV 617  
QY 652 VSAHACYIDRGFFSDPTQWTAFLGLHDGQSRAPQVQERRLKRIISHPFNDFTPDYD 711  
Db 618 VTAHACL-----EKFSPAIYKVVLGAM--QETRLERDVQIKGYTKMFLP-----YRAD 665  
QY 712 IALIELEKPAVSSMVRPTCLPDASHVFPAGKAIWVTGHTQYGTGALILQGEIRVI 771  
Db 666 IALLKLSSPAIIITDKHPACLPSNMYWADRLCYITGWEFTK--GYGAGILKEAQLPVI 724  
QY 772 NQTTCC--ENLLPQOITRRMVCVGLSGVDSGCGDGLSSVADRIIFAGVVMGSD 829  
Db 725 ENKYCNROSFLNGVNRSTELCAGHLAGVDSGQDSGLVCFERK--RYIIQGVTSWGLG 783  
QY 830 CAGNKKGVYTRLPLFRDWIK 851  
Db 784 CARLTRGVYRVSRVSLQD 805

## RESULT 29

A57096  
nudel protein precursor - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 27-Oct-1995 #sequence\_revision 27-Oct-1995 #text\_change 09-Jul-2004  
C:Accession: A57096  
R:Hong, C.C.; Haehnleto, C.  
Cell 82, 785-794, 1995  
A:Title: An unusual mosaic protein with a protease domain, encoded by the nudel gene, is  
A:Reference number: A57096, MUID:95401268; PMID:7671306  
A:Accession: A57096  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-2616 <HON>  
A:Cross-references: UNIPROT:P96159; GB:U29153; NID:9984320; PID:AAA83086.1; PID:9984321  
C:Genetic: A  
A:Gene: FlyBase:nd1  
A:Cross-references: FlyBase:FBgn0002926  
C:Superfamily: nudel protein; LDL receptor ligand-binding repeat homology; trypsin homol  
C:Keywords: extracellular protein  
F:891-929/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
F:1145-1378/Domain: trypsin homology <TRY>  
F:1196-1430/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
F:1776-1811/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
F:2310-2344/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
F:2351-2387/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
F:2421-2457/Domain: LDL receptor ligand-binding repeat homology <LDL3>

Query Match 10.0%; Score 467.5; DB 2; Length 2616;  
Best Local Similarity 22.9%; Pred. No. 3.3e-22;  
Matches 210; Conservative 104; Mismatches 308; Indels 295; Gaps 41;

QY 97 TNEFNVAVENS-----NSTEFVSLAKVNDALTLVSGVFLGPKYKESAVT 144  
Db 601 SSETVTTCYERQQLWKFFEDARPNQNELITGTALDAKLDDKG-PKIPPLN----- 653  
QY 145 AFSEGVYIAYVWSFSLPQHLVEEAERVMAEERVVMLEPPARSLKSFVVTSVVAFPTDSK 204  
Db 654 -----GHTWN-----AADAQILSLCEVAL---RMR-----NKVATMSDGE 686

QY 205 TYOR-----TQDNCSEFGLHARGVELMRFTTGPFPDPSYPAAHARQMALRGDADSLTF 260  
Db 687 TKKGETFTTSPSYQFTSRAPG-----GFPVSE-----TKKSAQCMFPNPF 729  
QY 261 RSPDLASCDERGDLYVTYNTLSPMEPHALVOLGTYPPSYNTLPHSSQNLVLTITNT 320  
Db 730 GMSIPVC-----FYMTPAIFRPMWMT-----PTMKGQAHFGSSN----- 769  
QY 321 ERHPRGEATFPQLPRN-----SSCGRLRKAGTFNSPYTYGHYPRNIDCT 367  
Db 770 ----PG--AGIFVPPQFGPSGNFPFGSSGSGAGGQGANIFSKNASQKPTNQOQOYCS 823  
QY 368 WNIENVNNGVKKRFKFFYLLBPQVPACTPKDQVVEINGEKCYGERSQFVTSNSKITY 427  
Db 824 Y---MNO-----SGRGAGS-----QTSQQQGGGSAF---SNAN-FKM 857  
QY 428 RPHSDQSYTD-----TGFLAEYLSYDSDPCPGQFTC--RTGRCIRKELRCDMA 475  
Db 858 R-HANQSTANQOQIIVASYAGLPQPIQERSRCPEPDQFSCGQGECLPAARMCDNV 916  
QY 476 DCTDHSBELNCSCDAGHFTCKKFKCPLFWVCDVNDCCGDNDEQGC-SCPAOTFPC-- 532  
Db 917 DCSGSDSAC-----TCADRVDER--LCDGYEDCPMGDEBLCFCGCESLAYSCE 966  
QY 533 -----SNKCLSKSQOCNGKDDGSDSDEASCPRKVVYCTK----- 569  
Db 967 NPQDFAKRNSTISMCTSRLEKCDFMNCLANGDEBECN--LVTVDADHSHGASASEG 1024  
QY 570 ---HTYR-----CLNG-----LCLSKNPECDGKEDCS----- 594  
Db 1025 YIVHNYRGDWHPCNNNGEKMAALACQMDENSRDHASLNVSTQTLTLPPTFIEPSLHAG 1084  
QY 595 -----DSDEKD-----CDGCL-----BSFTQAR 614  
Db 1085 VHFQAQCHGRNHSIDLVDHVAAYKCPMQGCLPSKSMLESHSKVRRAVDSKRIYDGR 1144  
QY 615 VGGTDADGEMPMOVSLHALGOGHICGASLISPNLVSAHACYIDRGFRYSDPTQWTA 674  
Db 1145 IVGSHTSALQMPVVAIVYNGKFN--CGGTYISDRMTISAAHCYINQKTFYE-----V 1197  
QY 675 FLGLHDGQSRAPQVQERRLKRIISHPFNDFTPDYDIALIELEKPAVSSMVRPTCLPD 734  
Db 1198 RAGILRRSSYS--PATQIQPVSHVYHQAYSRRSMBNDLSLRLNPLQFRMVRVPTCLPD 1256  
QY 735 ASHVFPAGKAIW-----VTGWHGTQYGTGALILQNG-----ELRVINQTTGEN 778  
Db 1257 KGRITVGDWIMQVBEHTLCTVVGW-----ALREKGPSSDPNROVIVPIRKCTD 1307  
QY 779 LLPQOITRRMVCVGLSGVDSGCGDGL--SYVADGRIFGAGVVMGDCAGRNKP 836  
Db 1308 --PEDQASEDICAGDPGGRDACQGDGGPLFCRSYSNADDFYLAGVSHGNGCARQER 1365  
QY 837 GVTYTRLPLFRDWIKENT 853  
Db 1366 GVTYTRVTLVLDWLEMAT 1382

RESULT 30  
A30351  
coagulation factor IXa (BC 3.4.21.22) precursor - dog  
C:Species: Canis lupus familiaris (dog)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: A30351; I46201  
R:Evans, J.P.; Walzke, H.H.; Ware, J.L.; Stafford, D.W.; High, K.A.  
Blood 74, 207-212, 1989  
A:Title: Molecular cloning of a cDNA encoding canine factor IX.  
A:Reference number: A30351; MUID:89323338; PMID:2752110  
A:Accession: A30351  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-452 <EVA>  
A:Cross-references: UNIPROT:P19540; GB:M21757; NID:9792719; PID:AAA75006.1; PID:9163948

Query Match 10.0%; Score 467.5; DB 2; Length 2616;  
Best Local Similarity 22.9%; Pred. No. 3.3e-22;  
Matches 210; Conservative 104; Mismatches 308; Indels 295; Gaps 41;

R; Axelrod, J.H.; Read, M.S.; Brinkhous, K.M.; Verma, I.M.  
Proc. Natl. Acad. Sci. U.S.A. 87, 5173-5177, 1990

Proc. Natl. Acad. Sci. U.S.A. 87, 5173-5177, 1990

A/Title: Phenotypic correction of factor IX deficiency  
A/Reference number: I46201; MUID:90311364; PMID:2367529

A/Accession: I46201  
A/Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA  
A;Residues: 1-452 <AXE>

A;Residues: 1-452 <AXE>

A: Cross-references: GB:M33826; NID:gl63949; PIDN:AAA30844.1; PID:gl639500  
C: Superfamily: coagulation factor X; EGF homology; Gla domain homology;

**C/Keywords:** beta-hydroxyaspartic acid, blood coagulation; calcium binding; carboxylglutamate  
**F1-21/Domain:** signal sequence #status predicted <SIG>

```
F;1-21/Domain: signal sequence #status predicted <SIG>
```

F;22-40/Domain: propeptide #status predicted <PRO>  
F;24-84/Domain: Gla domain homology <GLA>

F:24-84/Domain: GLA domain homology <GLA>predic

F;41-452/Product: coagulation factor IX #status predicted <MAT>  
F;90-121/Domain: EGF homology <EG1>

F:90-121/Domain: EGF homology <EGF>

F;127-163/Domain: EGF homology <EG2>  
F;218-445/Domain: trypsin homology <

E:218-445/Domain: EGF homology <EGZ>  
F:127-183/Domain: EGF homology <EGZ>

Modified site: gamma-carboxyglutamic acid (Glu) #8  
F:46,47,54,56,59,60,65,66,69,72,75,79

F: 46, 47, 54, 56, 59, 60, 65, 66, 69, 72, 75, 79/Modified site: gamma-carboxyglutamic acid (Glu) #88  
E: 57-62, 90-101, 95-110, 112-121, 127-138, 134-148, 150-163, 171-326, 243-259, 373-387, 398-426/Disordered

F;258,306,402/Active site: His, Asp, Ser #status predicted

Query Match	Score	DB 1;	Length
10.0%;	466.5;	DB 1;	452;
28.5%;	10.2;	DB 1;	452;

Best Local Similarity 28.5%; Freq. NO. 5,18-23;  
Matches 140; Conservative 63; Mismatches 164; Indels 125; Gaps 19;

```

QY      359 KDYEINGEKYC-GERSQ-----VYTSNKKITVRPHSDSYTDTGFLAEYLSYDSDSPC 453
Db      46  EEFPRGNLTRECIIEKSCFEAREVFENETK-----TTEFWKQYVDGQO----- 89

```

```
Db      46 EEFVRGNLERECIEKCSFEAREVFENTEK-----TTFWKQYVDGDQ----- 89
```

Oy 454 PGOFTCRIGRCIRKELRCDGMADCTDHSDELNCSGDAGHQFTCXNKFCKBLEFWWCDSVND 513  
 Db 90 -----CESNPPCLN-----DG--YKDDINSIEYCEWRAGF----- 116

```

Db 90 -----CESNPNCLN-----DG--VCKKDINSYECWCRAGF----- 11

```

OY	514	CGDSDDEOCSCSPAQ-TFRCNSGKCLSKSQOCNGKDDCGGSDDEASCPRKVVVATCTKHYY	572
DB	117	-----EGNCELQDVTGNIKNKGC-----KQFKLQAPD-----NKVYCS	149

```
Db      117 -----EGKNC ELDVT CNIKNGRC-----KOECKLGPD-----NKVVC S----- 14
```

[illegible]

150 -СТТСХОЛ ЯЕНОБСЕРДВРНОСВНУНТСМТБРАЕПТ ЕСИМДВЕНСТЕПЕХТИ ДНУ? 20

606 L R S T T R O A R V V G G T D A D E G E M P M O V S L A H G G C H I C G A S L I S P M W L V S A A H C Y I D D R G F R 665

[illegible]

```

DB      205  TQPLNDFETVVGQDKAPGPF PMWOLLNGVDA - FCGGSLINEKVVIVAHCIPED --- 263
OY      666  YSDPTQMTAFGLIHDOQRSAAPGVQERRLRIISHPFEND--TFDYDIALLELEKPAEY 723

```

[illegible]

Db 264 ----VKITVAGEHNTKEKREHT -EOKRNVIRITIIHHSYNAVTINKKNHDIALLEDEPLTL 318  
QY 724 SSMWRPICLPDA--SHVFPAKALVWTGWTGYGTGALILQKEIRVINTQTCENLLP 781

22 1 23 03MWF1CUEDA--SHVFFRQKAIWIGWGH1YIGIGALILQNGEIKVINQI1LENDDF / 0

DB 349 NSVYTPICIAADREYSNIFLKFGSGYVSGMGVFNPKGRSASIIQLKVLPLYDRAATCLRSTK 378  
QY 782 QQLTPRMKCVGLSGVDSQQGDSGGPLSSVEAQRIFAGVYVSMGDCAGRNRPQVYR 841

782 Q01PRMVCVGLSGVDSCQGDSEGPLSSVEADGRIFGAGVSWMGDCAGRNKPEVYTR 84

Db	379	FTIYNMFCAGFHEGGKDS	CGDSGP	HYTEVEGISFLTGII	ISMGECCAMKKGITIK	437
QY	842	LPLFRDMIKENT	853			

QY 842 LPLFRDWIKENT 853

Search completed: November 29, 2004, 08:33:30  
Job time : 39.6873 secs

Job time : 39.6873 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:32:39 ; Search time 111.183 Seconds  
(without alignments)  
2727.603 Million cell updates/sec

Title: US-09-936-333-27

Perfect score: 4683  
Sequence: 1 MGSDBARKGGGCKPFGAGL.....PGVYTRLPFRDWIKENTGV 855

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1575965 seqs, 354694765 residues

Total number of hits satisfying chosen parameters: 1575965

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	4659	99.5	855 14 US-10-295-027-1185	Sequence 1185, App
2	4659	99.5	855 15 US-10-072-012-353	Sequence 353, App
3	4659	99.5	855 15 US-10-072-012-412	Sequence 412, App
4	4659	99.5	855 15 US-10-072-012-419	Sequence 419, App
5	4655	99.4	855 15 US-10-072-012-354	Sequence 354, App
6	4655	99.4	855 15 US-10-072-012-420	Sequence 420, App
7	4655	99.4	855 15 US-10-037-012-132	Sequence 132, App
8	4653	99.4	855 10 US-09-776-191-12	Sequence 2, App1
9	4653	99.4	855 14 US-10-099-700A-2	Sequence 2, App11
10	4653	99.4	855 14 US-10-190-030B-2	Sequence 2, App11
11	4653	99.4	855 14 US-10-302-840A-2	Sequence 2, App11
12	4653	99.4	855 14 US-10-267-219-2	Sequence 2, App11
13	4653	99.4	855 14 US-10-112-221A-2	Sequence 2, App11

14	4653	99.4	855 14 US-10-104-271-2	Sequence 2, App11
15	4653	99.4	855 14 US-10-147-211A-2	Sequence 2, App11
16	4653	99.4	855 15 US-10-156-214A-2	Sequence 2, App11
17	4653	99.4	855 15 US-10-072-012-352	Sequence 352, App
18	4653	99.4	855 15 US-10-072-012-411	Sequence 411, App
19	4653	99.4	855 15 US-10-072-012-418	Sequence 418, App
20	4653	99.4	855 15 US-10-600-187-2	Sequence 2, App11
21	4614	98.5	851 15 US-10-276-774-1798	Sequence 1798, App
22	4614	98.5	851 15 US-10-296-113-1143	Sequence 1143, App
23	4167	89.0	762 16 US-10-729-807-1	Sequence 11, App1
24	4158.5	88.8	782 14 US-10-097-340-312	Sequence 312, App
25	4092	87.4	757 15 US-10-072-012-44	Sequence 44, App1
26	3884	82.9	855 9 US-09-900-751-2	Sequence 2, App11
27	3884	82.9	855 15 US-10-072-012-355	Sequence 355, App
28	3884	82.6	855 15 US-10-072-012-413	Sequence 413, App
29	3866	82.6	855 15 US-10-072-012-356	Sequence 356, App
30	3866	82.6	855 15 US-10-072-012-414	Sequence 414, App
31	3866	82.6	855 15 US-10-072-012-417	Sequence 417, App
32	3793	81.0	902 15 US-10-333-743-3	Sequence 3, App11
33	3793	81.0	902 15 US-10-600-187-10	Sequence 10, App1
34	3793	81.0	902 16 US-10-297-987B-11	Sequence 11, App1
35	2986	63.8	620 9 US-09-925-301-1193	Sequence 1193, App
36	2660	56.8	845 15 US-10-072-012-415	Sequence 415, App
37	1305	27.9	241 10 US-09-776-191-50	Sequence 50, App1
38	1305	27.9	241 14 US-10-099-700A-4	Sequence 4, App11
39	1305	27.9	241 14 US-10-092-004A-2	Sequence 2, App11
40	1305	27.9	241 14 US-10-190-030B-4	Sequence 4, App11
41	1305	27.9	241 14 US-10-302-840A-4	Sequence 4, App11
42	1305	27.9	241 14 US-10-267-219-4	Sequence 4, App11
43	1305	27.9	241 14 US-10-112-221A-4	Sequence 4, App11
44	1305	27.9	241 14 US-10-104-271-4	Sequence 4, App11
45	1305	27.9	241 14 US-10-147-211A-4	Sequence 4, App11
46	1114.5	23.8	799 15 US-10-072-012-410	Sequence 410, App
47	1114.5	23.8	799 15 US-10-072-012-416	Sequence 416, App
48	1112	23.7	802 9 US-09-888-615-113	Sequence 113, App
49	1112	23.7	802 9 US-09-978-295A-169	Sequence 169, App
50	1112	23.7	802 9 US-09-978-697-169	Sequence 169, App
51	1112	23.7	802 9 US-09-978-192A-169	Sequence 169, App
52	1112	23.7	802 9 US-09-999-832A-169	Sequence 169, App
53	1112	23.7	802 10 US-09-978-189-169	Sequence 169, App
54	1112	23.7	802 10 US-09-978-608A-169	Sequence 169, App
55	1112	23.7	802 10 US-09-978-585A-169	Sequence 169, App
56	1112	23.7	802 10 US-09-978-193A-169	Sequence 169, App
57	1112	23.7	802 10 US-09-978-403A-169	Sequence 169, App
58	1112	23.7	802 10 US-09-978-564A-169	Sequence 169, App
59	1112	23.7	802 10 US-09-999-833A-169	Sequence 169, App
60	1112	23.7	802 10 US-09-981-915A-169	Sequence 169, App
61	1112	23.7	802 10 US-09-978-824-169	Sequence 169, App
62	1112	23.7	802 10 US-09-918-585A-169	Sequence 169, App
63	1112	23.7	802 10 US-09-999-834A-169	Sequence 169, App
64	1112	23.7	802 10 US-09-978-423A-169	Sequence 169, App
65	1112	23.7	802 10 US-09-978-193A-169	Sequence 169, App
66	1112	23.7	802 10 US-09-999-830A-169	Sequence 169, App
67	1112	23.7	802 10 US-09-978-157A-169	Sequence 169, App
68	1112	23.7	802 10 US-09-978-187B-169	Sequence 169, App
69	1112	23.7	802 10 US-09-978-643A-169	Sequence 169, App
70	1112	23.7	802 10 US-09-978-375A-169	Sequence 169, App
71	1112	23.7	802 10 US-09-978-298A-169	Sequence 169, App
72	1112	23.7	802 10 US-09-978-681A-169	Sequence 169, App
73	1112	23.7	802 10 US-09-978-188A-169	Sequence 169, App
74	1112	23.7	802 10 US-09-978-194A-169	Sequence 169, App
75	1112	23.7	802 10 US-09-999-829A-169	Sequence 169, App
76	1112	23.7	802 10 US-09-978-299A-169	Sequence 169, App
77	1112	23.7	802 10 US-09-978-544A-169	Sequence 169, App
78	1112	23.7	802 10 US-09-978-802A-169	Sequence 169, App
79	1112	23.7	802 10 US-09-978-805A-169	Sequence 169, App
80	1112	23.7	802 11 US-09-999-831A-169	Sequence 169, App
81	1112	23.7	802 14 US-10-017-081A-169	Sequence 169, App
82	1112	23.7	802 14 US-10-167-749-169	Sequence 169, App
83	1112	23.7	802 14 US-10-013-921A-169	Sequence 169, App
84	1112	23.7	802 14 US-10-013-929A-169	Sequence 169, App
85	1112	23.7	802 14 US-10-016-177A-169	Sequence 169, App
86	1112	23.7	802 14 US-10-166-709A-169	Sequence 169, App

```
87 1112 23.7 802 14 US-10-143-031A-169 Sequence 169, App
88 1112 23.7 802 14 US-10-143-030A-169 Sequence 169, App
89 1112 23.7 802 14 US-10-002-967A-169 Sequence 169, App
90 1112 23.7 802 14 US-10-017-083A-169 Sequence 169, App
91 1112 23.7 802 14 US-10-145-128A-169 Sequence 169, App
92 1112 23.7 802 14 US-10-017-191A-169 Sequence 169, App
93 1112 23.7 802 14 US-10-143-028A-169 Sequence 169, App
94 1112 23.7 802 14 US-10-143-029A-169 Sequence 169, App
95 1112 23.7 802 14 US-10-145-089A-169 Sequence 169, App
96 1112 23.7 802 14 US-10-165-067A-169 Sequence 169, App
97 1112 23.7 802 14 US-10-145-017A-169 Sequence 169, App
98 1112 23.7 802 14 US-10-164-728A-169 Sequence 169, App
99 1112 23.7 802 14 US-10-013-926A-169 Sequence 169, App
100 1112 23.7 802 14 US-10-165-247A-169 Sequence 169, App
```

## ALIGNMENTS

```
RESULT 1
US-10-295-027-1185
; Sequence 1185, Application US/10295027
; Publication No. US2003023350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Nacasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevez, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OR INVENTION: Methods of diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US/10/295, 027
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/663, 733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350, 666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335, 394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332, 464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334, 393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340, 376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347, 211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347, 349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355, 250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356, 714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1185
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1185
```

```
Query Match 99.5%; Score 4659; DB 14; Length 855;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

1 MGSDDRARKGGGKPKDAGALKYNSRHEKVGLEEGVEFLPVNNVKKVEKKGPRWVLLAA 60

```
|||||
Db 1 MGSDDRARKGGGKPKDAGALKYNSRHEKVGLEEGVEFLPVNNVKKVEKKGPRWVLLAA 60
61 VILGILLVILGIGELVWHLQYRDPYRVOYKNGVMRTTNENFVDAYENNSTEVSLSKY 120
61 VILGILLVILGIGELVWHLQYRDPYRVOYKNGVMRTTNENFVDAYENNSTEVSLSKY 120
61 VILGILLVILGIGELVWHLQYRDPYRVOYKNGVMRTTNENFVDAYENNSTEVSLSKY 120
121 KDALKLISGVPPLGPHKESAVTAESGSLVIAYYSEFSIPQHLVEBAERVAAERVVM 180
121 KDALKLISGVPPLGPHKESAVTAESGSLVIAYYSEFSIPQHLVEBAERVAAERVVM 180
121 KDALKLISGVPPLGPHKESAVTAESGSLVIAYYSEFSIPQHLVEBAERVAAERVVM 180
181 LPPRASLSKSFVTVSVVAPPTDSKTYQRTQDNCSCGLHARGVELMRFTTPGPPDPYPA 240
181 LPPRASLSKSFVTVSVVAPPTDSKTYQRTQDNCSCGLHARGVELMRFTTPGPPDPYPA 240
181 LPPRASLSKSFVTVSVVAPPTDSKTYQRTQDNCSCGLHARGVELMRFTTPGPPDPYPA 240
241 HARCQNALRGDADSVLTPFRSFDLASCBERSGDLVTVNTVLSPMEPHLVOLCGTYPYS 300
241 HARCQNALRGDADSVLTPFRSFDLASCBERSGDLVTVNTVLSPMEPHLVOLCGTYPYS 300
301 YNLTFHSSQNVLLITLITNTERRHGFEATFFQLPRMSSCGRLRAOGTFNPGHY 360
301 YNLTFHSSQNVLLITLITNTERRHGFEATFFQLPRMSSCGRLRAOGTFNPGHY 360
361 PENIDCTWNI EVDNNOHVKRRFFPYLLEBPVPACTCPDYVAINBKTCGERSQFVYVS 420
361 PENIDCTWNI EVDNNOHVKRRFFPYLLEBPVPACTCPDYVAINBKTCGERSQFVYVS 420
421 NSNKITVRHSDSYNDTGFIAEYLSYSSDPDPCPGQFTCRGCIKRELCDMACTPH 480
421 NSNKITVRHSDSYNDTGFIAEYLSYSSDPDPCPGQFTCRGCIKRELCDMACTPH 480
481 SDELINSCDAGHOFCKNKECKPLFWVCDSVNDGDSNDEQSCSPAQTFRCNGKCLSK 540
481 SDELINSCDAGHOFCKNKECKPLFWVCDSVNDGDSNDEQSCSPAQTFRCNGKCLSK 540
541 SQQCNKDCDGDSDASCPKXNVVTCTKTYTRCLNGLCLSKNPECDGKEDSDSDSK 600
541 SQQCNKDCDGDSDASCPKXNVVTCTKTYTRCLNGLCLSKNPECDGKEDSDSDSK 600
601 DCCGGLRSFTROARVVGTDADGEMPMOVSJLHALGQHCASLISPMVLVSAACIYD 660
601 DCCGGLRSFTROARVVGTDADGEMPMOVSJLHALGQHCASLISPMVLVSAACIYD 660
661 DRGFRYSDFPTOWTAFLGLHDOSGRSAPGVQERLKR1ISHPFNDFTFYDIALLEKX 720
661 DRGFRYSDFPTOWTAFLGLHDOSGRSAPGVQERLKR1ISHPFNDFTFYDIALLEKX 720
721 AEYSSWVRPCLCPDASHVFPAGKAIWVTGHTQYGGTALILQKEIRIVINOTGENLL 780
721 AEYSSWVRPCLCPDASHVFPAGKAIWVTGHTQYGGTALILQKEIRIVINOTGENLL 780
781 PQGITERMMKCVGLSGGVDSGQSDSGPLSSVEADGRIFGAGVVSMDGCAQNNKPGVYT 840
781 PQGITERMMKCVGLSGGVDSGQSDSGPLSSVEADGRIFGAGVVSMDGCAQNNKPGVYT 840
841 RLPLFRDWIKENTGV 855
841 RLPLFRDWIKENTGV 855
```

```
RESULT 2
US-10-072-012-353
; Sequence 353, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Splet, Kimberly
; APPLICANT: Zethusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shimkete, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Bsha
; APPLICANT: Padigar, Muralidhara
```

APPLICANT: Anderson, David W.  
APPLICANT: Rastelli, Luca  
APPLICANT: Miller, Charles E.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Taupier Jr, Raymond J.  
APPLICANT: Gusev, Vladimir Y.  
APPLICANT: Coleman, Steven D.  
APPLICANT: Wolenc, Adam R.  
APPLICANT: Pena, Carol E. A  
APPLICANT: Furtak, Katarzyna  
APPLICANT: Grose, William M.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Lepley, Denise M.  
APPLICANT: Rieger, Daniel K.  
APPLICANT: Burgess, Catherine E.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-258  
CURRENT APPLICATION NUMBER: US/10/072.012  
CURRENT FILING DATE: 2002-01-31  
PRIOR APPLICATION NUMBER: 60/265,102  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: 60/265,514  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,517  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,412  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,395  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/266,406  
PRIOR FILING DATE: 2001-02-02  
PRIOR APPLICATION NUMBER: 60/266,767  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: 60/267,057  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/266,975  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/267,459  
PRIOR FILING DATE: 2001-02-08  
Remaining Prior Application data removed - See file wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1391  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 353  
LENGTH: 855  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-072-012-353

Query Match 99.5%; Score 4659; DB 15; Length 855;  
Best Local Similarity 99.6%; Pred. No. 0; Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSRRARKGGGPGYDFAGLKYNSRHEKVNGLIEGVEFLPVNNVYKVKHGPGRWVLA 60  
DB 1 MGSRRARKGGGPGYDFAGLKYNSRHEKVNGLIEGVEFLPVNNVYKVKHGPGRWVLA 60

QY 61 VLIGLLVLLGLIGLVWHLQYRDVAVQYKNGYKRYITENVDAYENSNSTEFVSLASKV 120  
DB 61 VLIGLLVLLGLIGLVWHLQYRDVAVQYKNGYKRYITENVDAYENSNSTEFVSLASKV 120

QY 121 KDALKLTVSGVPFLGIPYHESAVTAFSGSVIAYYWSFSLPOHLVEAEKVMABEERVVM 180  
DB 121 KDALKLTVSGVPFLGIPYHESAVTAFSGSVIAYYWSFSLPOHLVEAEKVMABEERVVM 180

QY 181 LPPRARSLSKSVVAVFPTDSKTVQRTQDNCSCFGLHARGVLEMRFTTGPDPSPYPA 240  
DB 181 LPPRARSLSKSVVAVFPTDSKTVQRTQDNCSCFGLHARGVLEMRFTTGPDPSPYPA 240

QY 241 HARCOWALRGADSVLSTFRSFDLASCDEGSDLVYTYNTLSPEHPALVQLGCTYPPS 300  
DB 241 HARCOWALRGADSVLSTFRSFDLASCDEGSDLVYTYNTLSPEHPALVQLGCTYPPS 300

QY 301 YNLFFHSQVLLITLLINTERBRHGFATTFQLPKRMSSCGRLRKAQGTFSPPYPGHY 360

DB 301 YNLFFHSQVLLITLLINTERBRHGFATTFQLPKRMSSCGRLRKAQGTFSPPYPGHY 360  
QY 361 PPNIIDCTWNIIEVPNNQVAVKRFKFFYLLEPGVAGTCKPYVLENGKCYGESSQFVYTS 420  
DB 361 PPNIIDCTWNIIEVPNNQVAVKRFKFFYLLEPGVAGTCKPYVLENGKCYGESSQFVYTS 420

QY 421 NSNKITVRFPSDOSYTDGTGLAEYLSYDSSDPGCGQFTCRGRCIRKELACDGMADCTDH 480  
DB 421 NSNKITVRFPSDOSYTDGTGLAEYLSYDSSDPGCGQFTCRGRCIRKELACDGMADCTDH 480

QY 481 SDELNCSGDAHQFTCKNKECKPLFWVCDSVNDGDNDSDEGSCSPAQTFRCNGKCLSK 540  
DB 481 SDELNCSGDAHQFTCKNKECKPLFWVCDSVNDGDNDSDEGSCSPAQTFRCNGKCLSK 540

QY 541 SQQCNKGDDGSDSDSACRNVAVTCTKTYRCLNCLCSKGNPECDGDESDGSDGDEK 600  
DB 541 SQQCNKGDDGSDSDSACRNVAVTCTKTYRCLNCLCSKGNPECDGDESDGSDGDEK 600

QY 601 DDCGRLSFTROARVVGCTDADGEMPMOYSLHALGQGHICGASLISPMNLVSAHACYID 660  
DB 601 DDCGRLSFTROARVVGCTDADGEMPMOYSLHALGQGHICGASLISPMNLVSAHACYID 660

QY 661 DRGFRYSDPQWTAFLGLHDSQSRASAPGVERRLKRILISHPFNDFTFDYDIALLEKXP 720  
DB 661 DRGFRYSDPQWTAFLGLHDSQSRASAPGVERRLKRILISHPFNDFTFDYDIALLEKXP 720

QY 721 AEVSSMVRPITCLDPASVFPAGKAIWYTGWHGYOYGGTGLIIOKGBIRIYNOTTCNLL 780  
DB 721 AEVSSMVRPITCLDPASVFPAGKAIWYTGWHGYOYGGTGLIIOKGBIRIYNOTTCNLL 780

QY 781 PQOITPRMCMVGFISGSDVSCGSDGGLSSVEADGRIFPAGVYVWMDGACGRNKPQVYT 840  
DB 781 PQOITPRMCMVGFISGSDVSCGSDGGLSSVEADGRIFPAGVYVWMDGACGRNKPQVYT 840

QY 841 RLPLFRDMIKENTGV 855  
DB 841 RLPLFRDMIKENTGV 855

RESULT 3  
US-10-072-012-412  
Sequence 412, Application US/10072012  
Publication No. US20040033493A1  
GENERAL INFORMATION:  
APPLICANT: Tchiernev, Velizar  
APPLICANT: Spytek, Kimberly  
APPLICANT: Zernusen, Bryan  
APPLICANT: Patuturajan, Meera  
APPLICANT: Shimkets, Richard  
APPLICANT: Li, Li  
APPLICANT: Gangoli, Baha  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Anderson, David W.  
APPLICANT: Rastelli, Luca  
APPLICANT: Miller, Charles E.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Taupier Jr, Raymond J.  
APPLICANT: Gusev, Vladimir Y.  
APPLICANT: Coleman, Steven D.  
APPLICANT: Wolenc, Adam R.  
APPLICANT: Pena, Carol E. A  
APPLICANT: Furtak, Katarzyna  
APPLICANT: Grose, William M.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Lepley, Denise M.  
APPLICANT: Rieger, Daniel K.  
APPLICANT: Burgess, Catherine E.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-258  
CURRENT APPLICATION NUMBER: US/10/072.012  
CURRENT FILING DATE: 2002-01-31  
PRIOR APPLICATION NUMBER: 60/265,102



```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 412
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-412

```

```

Query Match 99.5%; Score 4659; DB 15; Length 855;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 MGSDRAKGGGPGDAGAGLKYNSRHKVNGLEBGEVFLPVNNVKKYKKGPRGMVLLAA 60
DB 1 MGSDRAKGGGPGDAGAGLKYNSRHKVNGLEBGEVFLPVNNVKKYKKGPRGMVLLAA 60
QY 61 VLIGLLVLIGLIGFLVHLQYRDVAVQKXKGYRITNENFVDAYENSSTEFVSLASKV 120
DB 61 VLIGLLVLIGLIGFLVHLQYRDVAVQKXKGYRITNENFVDAYENSSTEFVSLASKV 120
QY 121 KDLAKLISYGVPLGPRYHKSAAVAFSEGSVIAIYMWSEFSPQHLVEARVMAEERVVM 180
DB 121 KDLAKLISYGVPLGPRYHKSAAVAFSEGSVIAIYMWSEFSPQHLVEARVMAEERVVM 180
QY 181 LPPRARSLSKFFVYVVAFAFPTDSKTVORTDNSSCFGLHARGVLMKFTTGGFPPDSYPA 240
DB 181 LPPRARSLSKFFVYVVAFAFPTDSKTVORTDNSSCFGLHARGVLMKFTTGGFPPDSYPA 240
QY 241 HARCQWALRGDADSVLSLTFRSPDLASCDESGSDLVTVYNTLSFMEPHALVQLCGTYP 300
DB 241 HARCQWALRGDADSVLSLTFRSPDLASCDESGSDLVTVYNTLSFMEPHALVQLCGTYP 300
QY 301 YNLFHSSQNVLLITLTNTERHRRHGFEPATFQOLPRMSGGGRIRKQGFNPSYVYGHY 360
DB 301 YNLFHSSQNVLLITLTNTERHRRHGFEPATFQOLPRMSGGGRIRKQGFNPSYVYGHY 360
QY 361 PPNIDCTWNTLVNNOHVAVRKFKEFYLLBPGVAPGCTPKDQVBEINSGYSGERSQFVYTS 420
DB 361 PPNIDCTWNTLVNNOHVAVRKFKEFYLLBPGVAPGCTPKDQVBEINSGYSGERSQFVYTS 420
QY 421 NSNKTITVRFHSDSYDTGTGLAEYLSYDSDDPCGQGTCTGRCIRKELRCDGADCTDH 480
DB 421 NSNKTITVRFHSDSYDTGTGLAEYLSYDSDDPCGQGTCTGRCIRKELRCDGADCTDH 480
QY 481 SDELNGSCDAGHQTCKNKKCPFLFWCDVNDGDGSDGSCCPAQTFRCSNGKLSK 540
DB 481 SDELNGSCDAGHQTCKNKKCPFLFWCDVNDGDGSDGSCCPAQTFRCSNGKLSK 540
QY 541 SQQCGKDDGSDGSDGSCCPVNVVCTKHTYRCLNGLCLSKNGPECDEKEDSDGSDG 600
DB 541 SQQCGKDDGSDGSDGSCCPVNVVCTKHTYRCLNGLCLSKNGPECDEKEDSDGSDG 600
QY 601 DCCGRLSFTRQARVVGCTADGSEMPQVSLHALGGHICGASLISPNMLVSAHACYID 660

```

```

DB 601 DCCGRLSFTRQARVVGCTADGSEMPQVSLHALGGHICGASLISPNMLVSAHACYID 660
QY 661 DRGFRYSDDPTQWAFPIGLHDSORSA PGVERLKRIIISHPFNDFTFDIDLLELEXP 720
DB 661 DRGFRYSDDPTQWAFPIGLHDSORSA PGVERLKRIIISHPFNDFTFDIDLLELEXP 720
QY 721 AEYSSWVRPICLDPAHVFPAGKAIWVTGNGHTQYGGTGLILQKGEIRVINTCCENLL 780
DB 721 AEYSSWVRPICLDPAHVFPAGKAIWVTGNGHTQYGGTGLILQKGEIRVINTCCENLL 780
QY 781 PQOITPRMVCVGLSGVDSCQDSDGGLSSVADRIIRAGVYVSGDGAGNKGVTY 840
DB 781 PQOITPRMVCVGLSGVDSCQDSDGGLSSVADRIIRAGVYVSGDGAGNKGVTY 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

```

```

RESULT 4
US-10-072-012-419
; Sequence 419, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tcheurev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zethusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkete, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Esha
; APPLICANT: Padigaru, Muraidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Futrak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; Remaining Prior Application data removed - See File Wrapper or PALM.

```

NUMBER OF SEQ ID NOS: 1391  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 419  
LENGTH: 855  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-072-012-419

Query Match 99.5%; Score 4659; DB 15; Length 855;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKNGLEGEVEFLPVNNVYKVEKHPGRWVTLAA 60
DB 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKNGLEGEVEFLPVNNVYKVEKHPGRWVTLAA 60
QY 61 VLIGLLVLLIGLIGFLVHLYQYRDVYQVKYNGYMRITNENFVDAYENSNSTEVSLSKV 120
DB 61 VLIGLLVLLIGLIGFLVHLYQYRDVYQVKYNGYMRITNENFVDAYENSNSTEVSLSKV 120
QY 121 KDALKLYSGVPFLGPHYKESAVTAFSGSIYAYWSEFSIPOHLYEAEKRVMAEERYVM 180
DB 121 KDALKLYSGVPFLGPHYKESAVTAFSGSIYAYWSEFSIPOHLYEAEKRVMAEERYVM 180
QY 181 LPPRARSLSKSVYVAVFPTDSKTVORTODNSCSFGLHARGVEIMRPTTGPDPSPYPA 240
DB 181 LPPRARSLSKSVYVAVFPTDSKTVORTODNSCSFGLHARGVEIMRPTTGPDPSPYPA 240
QY 241 HARCQMLRGDADSVLSTFRSFDLASCDERGSDLVTVNTLSPMEPHALVOLGCTYPPS 300
DB 241 HARCQMLRGDADSVLSTFRSFDLASCDERGSDLVTVNTLSPMEPHALVOLGCTYPPS 300
QY 301 YNLTFHSSQNLTLTLTNTERRHHPGFATFQLPKMSCCGRIRKAGSTNSPYRGHY 360
DB 301 YNLTFHSSQNLTLTLTNTERRHHPGFATFQLPKMSCCGRIRKAGSTNSPYRGHY 360
QY 361 PPNIDCTWNIENVNNQHVAFKFFYLLEPGVAGTCCKDYVEINGEYKCGERQFVYTS 420
DB 361 PPNIDCTWNIENVNNQHVAFKFFYLLEPGVAGTCCKDYVEINGEYKCGERQFVYTS 420
QY 421 NSNKITVRFHSDQSYDTGFLAEYLSYDSDPCPGQFTCRGRCIRKELCDGNADCTDH 480
DB 421 NSNKITVRFHSDQSYDTGFLAEYLSYDSDPCPGQFTCRGRCIRKELCDGNADCTDH 480
QY 481 SDELNCSGDADHQTCKRKKFKPLFWVCDVNDGDNDSDEGSCCPAQTFRCSNGKCLSK 540
DB 481 SDELNCSGDADHQTCKRKKFKPLFWVCDVNDGDNDSDEGSCCPAQTFRCSNGKCLSK 540
QY 541 SQQNGKDDCGDSDGDEASCPKVNVTCTKHTYRCINGLCLSKNPECDGKEDCGSDSEK 600
DB 541 SQQNGKDDCGDSDGDEASCPKVNVTCTKHTYRCINGLCLSKNPECDGKEDCGSDSEK 600
QY 601 DCDGSLSFTRQARVVGSTADDEGEMWQVSLHALGQHICGASLISBNMLVSAHAYCID 660
DB 601 DCDGSLSFTRQARVVGSTADDEGEMWQVSLHALGQHICGASLISBNMLVSAHAYCID 660
QY 661 DRGRYSIDPTQWTFGLGHDQSORSAPGVQERRLRKRIISHPFDFDYDIALLELEKP 720
DB 661 DRGRYSIDPTQWTFGLGHDQSORSAPGVQERRLRKRIISHPFDFDYDIALLELEKP 720
QY 721 AEYSMWVPICTLPASHVFPAGKAIWYGMHTQYGGGAILLKGELRVNNTTCENLL 780
DB 721 AEYSMWVPICTLPASHVFPAGKAIWYGMHTQYGGGAILLKGELRVNNTTCENLL 780
QY 781 PQQITPRMVCVFLSGVDSCQDGGGFLSSVEADGRIFGAGVYVWGDGAGRRKPGVYT 840
DB 781 PQQITPRMVCVFLSGVDSCQDGGGFLSSVEADGRIFGAGVYVWGDGAGRRKPGVYT 840
QY 841 RLPLFRDMIKENTGV 855
DB 841 RLPLFRDMIKENTGV 855
```

## RESULT 5

US-10-072-012-354

Sequence 354, Application US/10072012

Publication No. US2004003493A1

GENERAL INFORMATION:

APPLICANT: Tchenev, Velizar

APPLICANT: Zephusen, Bryan

APPLICANT: Zephusen, Bryan

APPLICANT: Shukets, Richard

APPLICANT: Li, Li

APPLICANT: Gangoli, Esna

APPLICANT: Padigaru, Muralidhara

APPLICANT: Anderson, David W.

APPLICANT: Raestell, Luca

APPLICANT: Miller, Charles E.

APPLICANT: Gerlach, Valerie

APPLICANT: Taupier Jr, Raymond J.

APPLICANT: Gueev, Vladimir Y.

APPLICANT: Coleman, Steven D.

APPLICANT: Wolenc, Adam R.

APPLICANT: Pena, Carol E. A

APPLICANT: Furtak, Katarzyna

APPLICANT: Grose, William M.

APPLICANT: Alsobrook II, John P.

APPLICANT: Lepley, Denise M.

APPLICANT: Rieger, Daniel K.

APPLICANT: Burgess, Catherine E.

TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-258

CURRENT APPLICATION NUMBER: US/10/072,012

CURRENT FILING DATE: 2002-01-31

PRIOR APPLICATION NUMBER: 60/265,102

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: 60/265,514

PRIOR FILING DATE: 2001-01-31

PRIOR APPLICATION NUMBER: 60/265,517

PRIOR FILING DATE: 2001-01-31

PRIOR APPLICATION NUMBER: 60/265,412

PRIOR FILING DATE: 2001-01-31

PRIOR APPLICATION NUMBER: 60/265,395

PRIOR FILING DATE: 2001-01-31

PRIOR APPLICATION NUMBER: 60/266,406

PRIOR FILING DATE: 2001-02-02

PRIOR APPLICATION NUMBER: 60/266,767

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: 60/267,057

PRIOR FILING DATE: 2001-02-07

PRIOR APPLICATION NUMBER: 60/266,975

PRIOR FILING DATE: 2001-02-07

PRIOR APPLICATION NUMBER: 60/267,459

PRIOR FILING DATE: 2001-02-08

REMAINING PRIOR Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 1391

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 354

LENGTH: 855

TYPE: PRT

ORGANISM: Homo sapiens

US-10-072-012-354

Query Match 99.4%; Score 4655; DB 15; Length 855;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKNGLEGEVEFLPVNNVYKVEKHPGRWVTLAA 60
DB 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKNGLEGEVEFLPVNNVYKVEKHPGRWVTLAA 60
QY 61 VLIGLLVLLIGLIGFLVHLYQYRDVYQVKYNGYMRITNENFVDAYENSNSTEVSLSKV 120
DB 61 VLIGLLVLLIGLIGFLVHLYQYRDVYQVKYNGYMRITNENFVDAYENSNSTEVSLSKV 120
```

```

QY 121 KQALKLLYSVPLGPGYHKSASVTAFASEGSAIAYYMSFESI PQHLYEAEERVAEERVM 180
Db 121 KQALKLLYSVPLGPGYHKSASVTAFASEGSAIAYYMSFESI PQHLYEAEERVAEERVM 180
QY 181 LPPRARSLSKSPVYTSVVAAPPTDSKTVOQTODNSCSFGLHARGVELMRFTTGPFPDSYPYA 240
Db 181 LPPRARSLSKSPVYTSVVAAPPTDSKTVOQTODNSCSFGLHARGVELMRFTTGPFPDSYPYA 240
QY 241 HARCQWALRGDADSVSLTFRSPDLASCDERGSGLVYVNTLSMPMEHALVOLCGTYPSPS 300
Db 241 HARCQWALRGDADSVSLTFRSPDLASCDERGSGLVYVNTLSMPMEHALVOLCGTYPSPS 300
QY 301 YNLTFHSSQNVLLITLITNTERRHHPGEATFPQLPRMSCGRLRKAQGTENSPYPYGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHHPGEATFPQLPRMSCGRLRKAQGTENSPYPYGHY 360
QY 361 PPNIDCTWNIIEVNNQHVAVRFKFFYLLEPGVPAGTCKPYVEINSEKTCGERSQFVYTS 420
Db 361 PPNIDCTWNIIEVNNQHVAVRFKFFYLLEPGVPAGTCKPYVEINSEKTCGERSQFVYTS 420
QY 421 NSNKITVRPHSDOSYTDYTGFLAEYLSYDSSDPCGQPTCRGRCIRKELRCDDMADCTDH 480
Db 421 NSNKITVRPHSDOSYTDYTGFLAEYLSYDSSDPCGQPTCRGRCIRKELRCDDMADCTDH 480
QY 481 SDEINCSGADGHOFTCKNKECKPLFWVCDSDVNDGSDNSDBQSCSPAQTFRCSNGKCLSK 540
Db 481 SDEINCSGADGHOFTCKNKECKPLFWVCDSDVNDGSDNSDBQSCSPAQTFRCSNGKCLSK 540
QY 541 SQQCNGDDDDGSDSDASCKRANVVTCTKATYRCLNLCISKNPEDCDGEDSDGSDSK 600
Db 541 SQQCNGDDDDGSDSDASCKRANVVTCTKATYRCLNLCISKNPEDCDGEDSDGSDSK 600
QY 601 DCCGGLSFTROARVVGSTDADEGEMPMQVSLHLAGGSHICGASLISPMNLVSAHICYID 660
Db 601 DCCGGLSFTROARVVGSTDADEGEMPMQVSLHLAGGSHICGASLISPMNLVSAHICYID 660
QY 661 DRGFRYSPTQWTAFLGLHDQSORAPGVQERLRKRIISHPFNDFTFDYDIALLELEKP 720
Db 661 DRGFRYSPTQWTAFLGLHDQSORAPGVQERLRKRIISHPFNDFTFDYDIALLELEKP 720
QY 721 AEYSSMWRTICLPDASHVPFAGKAIWTTGNGHTOYGGTGALLQKGIRIYNQTTGNNL 780
Db 721 AEYSSMWRTICLPDASHVPFAGKAIWTTGNGHTOYGGTGALLQKGIRIYNQTTGNNL 780
QY 781 PQQITPRMCMCVGLSGVDSCQSDGSGPLSSVEADGRIIFGAGVYVSWDGCAGRNKPPVYT 840
Db 781 PQQITPRMCMCVGLSGVDSCQSDGSGPLSSVEADGRIIFGAGVYVSWDGCAGRNKPPVYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

```

RESULT 6  
US-10-072-012-420  
Sequence 420, Application US/10072012  
Publication No. US2004003493A1  
GENERAL INFORMATION:

```

APPLICANT: Tchernev, Velizar
APPLICANT: Splytek, Kimberly
APPLICANT: Zernhuseen, Bryan
APPLICANT: Patturajan, Meera
APPLICANT: Shinkets, Richard
APPLICANT: Li, Li
APPLICANT: Gangoli, Esha
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles B.
APPLICANT: Gerlach, Valerie
APPLICANT: Tsaplier Jr, Raymond J.
APPLICANT: Gusev, Vladimyr Y.
APPLICANT: Colman, Steven D.

```

```

APPLICANT: Molenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsodrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072, 012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265, 102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265, 514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265, 517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265, 412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265, 395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266, 406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266, 767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267, 057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266, 975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267, 459
PRIOR FILING DATE: 2001-02-08
Remainder Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 420
LENGTH: 855
TYPE: PRT
ORGANISM: Homo sapiens
US-10-072-012-420

```

Query Match 99.4%; Score 4655; DB 15; Length 855;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 MSDBRARKGGGKQDAGLKYNSRHEKXNGLEBGEVFLPVNNVKKYERKGPQRWVLLA 60
Db 1 MSDBRARKGGGKQDAGLKYNSRHEKXNGLEBGEVFLPVNNVKKYERKGPQRWVLLA 60
QY 61 VILGLLVLLIGIGFLVWHLQYRDVRYQVKYNGYMRITNENFVDAYENSSTEFVSLASRY 120
Db 61 VILGLLVLLIGIGFLVWHLQYRDVRYQVKYNGYMRITNENFVDAYENSSTEFVSLASRY 120
QY 121 KQALKLLYSVPLGPGYHKSASVTAFASEGSAIAYYMSFESI PQHLYEAEERVAEERVM 180
Db 121 KQALKLLYSVPLGPGYHKSASVTAFASEGSAIAYYMSFESI PQHLYEAEERVAEERVM 180
QY 181 LPPRARSLSKSPVYTSVVAAPPTDSKTVOQTODNSCSFGLHARGVELMRFTTGPFPDSYPYA 240
Db 181 LPPRARSLSKSPVYTSVVAAPPTDSKTVOQTODNSCSFGLHARGVELMRFTTGPFPDSYPYA 240
QY 241 HARCQWALRGDADSVSLTFRSPDLASCDERGSGLVYVNTLSMPMEHALVOLCGTYPSPS 300
Db 241 HARCQWALRGDADSVSLTFRSPDLASCDERGSGLVYVNTLSMPMEHALVOLCGTYPSPS 300
QY 301 YNLTFHSSQNVLLITLITNTERRHHPGEATFPQLPRMSCGRLRKAQGTENSPYPYGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHHPGEATFPQLPRMSCGRLRKAQGTENSPYPYGHY 360
QY 361 PPNIDCTWNIIEVNNQHVAVRFKFFYLLEPGVPAGTCKPYVEINSEKTCGERSQFVYTS 420
Db 361 PPNIDCTWNIIEVNNQHVAVRFKFFYLLEPGVPAGTCKPYVEINSEKTCGERSQFVYTS 420

```

QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSDSDPCPGQFTCRGTGRCIRKELRCGMDACTDH 480  
Db 421 NSNKITVRFHSDQSYTDTGFLAEYLSDSDPCPGQFTCRGTGRCIRKELRCGMDACTDH 480  
QY 481 SDELINCSGDAHQFTCKKFKCKPLFWVCDSDVNDGSDNSDEGSCCPAQTFRCSNGKLSK 540  
Db 481 SDELINCSGDAHQFTCKKFKCKPLFWVCDSDVNDGSDNSDEGSCCPAQTFRCSNGKLSK 540  
QY 541 SQQNGKDDCGDSDGDEASCPVNVVVTCTKHTYRCLNGCLSKNGECCGKEDCDGSDGSK 600  
Db 541 SQQNGKDDCGDSDGDEASCPVNVVVTCTKHTYRCLNGCLSKNGECCGKEDCDGSDGSK 600  
QY 601 DCDGSLSFTRQARVVGSTDADEGEMPWQVSLHALGGHICGASLISPNMLVSAHACYD 660  
Db 601 DCDGSLSFTRQARVVGSTDADEGEMPWQVSLHALGGHICGASLISPNMLVSAHACYD 660  
QY 661 DRGRYSPTQWTFVLGLHDSQSRAPGVQERRRLRIISHPFNDFTFDYDIALLELEKP 720  
Db 661 DRGRYSPTQWTFVLGLHDSQSRAPGVQERRRLRIISHPFNDFTFDYDIALLELEKP 720  
QY 721 AEYSMWPRICLPDASHVFPAGKAIWVTGMHTGYCGTGALILQKGEIRVINOCTCENL 780  
Db 721 AEYSMWPRICLPDASHVFPAGKAIWVTGMHTGYCGTGALILQKGEIRVINOCTCENL 780  
QY 781 PQQITPRMVCVFLSGVSDSCGSDGSPLSVEADGRIFGAGVSWGDCAGRNKPGVYT 840  
Db 781 PQQITPRMVCVFLSGVSDSCGSDGSPLSVEADGRIFGAGVSWGDCAGRNKPGVYT 840  
QY 841 RLPLFRDWIKENTGV 855  
Db 841 RLPLFRDWIKENTGV 855

## RESULT 7

US-10-037-417-132  
Sequence 132, Application US/10037417  
Publication No. US20040052806A1  
GENERAL INFORMATION:  
APPLICANT: Kekuda, Rameeh  
APPLICANT: Alsobrook II, John P  
APPLICANT: Tcherenev, Veilizar T  
APPLICANT: Liu, Xiaohong  
APPLICANT: Spytek, Kimberly A  
APPLICANT: Paturejan, Meera  
APPLICANT: Groose, William M  
APPLICANT: Lepley, Denise M  
APPLICANT: Burgess, Catherine E  
APPLICANT: Verneer, Corine A.M.  
APPLICANT: Li, Li  
APPLICANT: Gorman, Linda  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Scioire, Paul  
APPLICANT: Ellerman, Karen  
APPLICANT: Malyankar, Uriel M  
APPLICANT: Rothenberg, Mark  
APPLICANT: Stone, David J  
APPLICANT: Boldog, Ferenc L  
APPLICANT: Guo, Xiaojia  
APPLICANT: Shenoy, Suresh G  
APPLICANT: Anderson, David W  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Taupier Jr, Raymond J  
APPLICANT: Miller, Charles E  
APPLICANT: Eisen, Andrew J  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-235  
CURRENT APPLICATION NUMBER: US/10/037,417  
CURRENT FILING DATE: 2002-09-20  
PRIOR APPLICATION NUMBER: 60/260,018  
PRIOR FILING DATE: 2001-01-05  
PRIOR APPLICATION NUMBER: 60/260,360  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: 60/272,411

QY 1 PRIOR FILING DATE: 2001-02-28  
Db 1 PRIOR APPLICATION NUMBER: 60/272,817  
QY 2 PRIOR FILING DATE: 2001-03-02  
Db 2 PRIOR APPLICATION NUMBER: 60/291,186  
QY 3 PRIOR FILING DATE: 2001-05-15  
Db 3 PRIOR APPLICATION NUMBER: 60/303,231  
QY 4 PRIOR FILING DATE: 2001-07-05  
Db 4 PRIOR APPLICATION NUMBER: 60/305,060  
QY 5 PRIOR FILING DATE: 2001-07-12  
Db 5 PRIOR APPLICATION NUMBER: 60/318,405  
QY 6 PRIOR FILING DATE: 2001-09-10  
Db 6 PRIOR APPLICATION NUMBER: 60/318,700  
QY 7 NUMBER OF SEQ ID NOS: 227  
Db 7 SOFTWARE: Patentin Ver. 2.1  
QY 8 SEQ ID NO 132  
Db 8 LENGTH: 855  
QY 9 TYPE: PRT  
Db 9 ORGANISM: Homo sapiens  
US-10-037-417-132

Query Match 99.4%; Score 4655; DB 15; Length 855;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSRRARKGGGPGKDFGAGLKYNRSRHEKNGLEGEVFLPVNNVKKYKKGPGRWVLLAA 60  
Db 1 MGSRRARKGGGPGKDFGAGLKYNRSRHEKNGLEGEVFLPVNNVKKYKKGPGRWVLLAA 60  
QY 61 VLIGLLVLLIGLGLVHMLQYRDVRYQKXNGYMRITNENFVDAEYNSSTFVSLASKV 120  
Db 61 VLIGLLVLLIGLGLVHMLQYRDVRYQKXNGYMRITNENFVDAEYNSSTFVSLASKV 120  
QY 121 KDALKLYSGVPLGFLPHKESAVTAPSEGSVIAIYMESESIPOHLYBEARVMAEERVVM 180  
Db 121 KDALKLYSGVPLGFLPHKESAVTAPSEGSVIAIYMESESIPOHLYBEARVMAEERVVM 180  
QY 121 KDALKLYSGVPLGFLPHKESAVTAPSEGSVIAIYMESESIPOHLYBEARVMAEERVVM 180  
Db 121 KDALKLYSGVPLGFLPHKESAVTAPSEGSVIAIYMESESIPOHLYBEARVMAEERVVM 180  
QY 181 LPPRARSLSKFVVTSSVVAFPDSTKVORTDONSCEFGLHARGVELMFTTGPFPDSDPYPA 240  
Db 181 LPPRARSLSKFVVTSSVVAFPDSTKVORTDONSCEFGLHARGVELMFTTGPFPDSDPYPA 240  
QY 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSDLVTVNTLSPEMHALVOLCGTYPSP 300  
Db 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSDLVTVNTLSPEMHALVOLCGTYPSP 300  
QY 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSDLVTVNTLSPEMHALVOLCGTYPSP 300  
Db 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSDLVTVNTLSPEMHALVOLCGTYPSP 300  
QY 301 YNLTFFSSQVNLITLITNTERRHPGEATFPQLPRMSGCCGRILRKAQGTFFNSPYRGHY 360  
Db 301 YNLTFFSSQVNLITLITNTERRHPGEATFPQLPRMSGCCGRILRKAQGTFFNSPYRGHY 360  
QY 361 PPNIDCTWNIENVPNQVAVKVPKFFYLEPGVAGTCPKDYVEINGEKYGERSQFVVTSS 420  
Db 361 PPNIDCTWNIENVPNQVAVKVPKFFYLEPGVAGTCPKDYVEINGEKYGERSQFVVTSS 420  
QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSDSDPCPGQFTCRGTGRCIRKELRCGMDACTDH 480  
Db 421 NSNKITVRFHSDQSYTDTGFLAEYLSDSDPCPGQFTCRGTGRCIRKELRCGMDACTDH 480  
QY 481 SDELINCSGDAHQFTCKKFKCKPLFWVCDSDVNDGSDNSDEGSCCPAQTFRCSNGKLSK 540  
Db 481 SDELINCSGDAHQFTCKKFKCKPLFWVCDSDVNDGSDNSDEGSCCPAQTFRCSNGKLSK 540  
QY 541 SQQNGKDDCGDSDGDEASCPVNVVVTCTKHTYRCLNGCLSKNGECCGKEDCDGSDGSK 600  
Db 541 SQQNGKDDCGDSDGDEASCPVNVVVTCTKHTYRCLNGCLSKNGECCGKEDCDGSDGSK 600  
QY 601 DCDGSLSFTRQARVVGSTDADEGEMPWQVSLHALGGHICGASLISPNMLVSAHACYD 660  
Db 601 DCDGSLSFTRQARVVGSTDADEGEMPWQVSLHALGGHICGASLISPNMLVSAHACYD 660  
QY 661 DRGRYSPTQWTFVLGLHDSQSRAPGVQERRRLRIISHPFNDFTFDYDIALLELEKP 720  
Db 661 DRGRYSPTQWTFVLGLHDSQSRAPGVQERRRLRIISHPFNDFTFDYDIALLELEKP 720

```
QY 721 AEYSWVRPCLPDASHVPAGKAIWVTGHTQYGGTALLI0KGEIRVINQTCENLL 780
DB 721 AEYSWVRPCLPDASHVPAGKAIWVTGHTQYGGTALLI0KGEIRVINQTCENLL 780
QY 781 PQQITPRMVCVGLSGVDSCQDSGGLPSVEADGRIFGAGVVSMDGCAQRNKPQVYT 840
DB 781 PQQITPRMVCVGLSGVDSCQDSGGLPSVEADGRIFGAGVVSMDGCAQRNKPQVYT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

RESULT 8
US-09-776-191-2
; Sequence 2, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiumn-Chern Yeh
; APPLICANT: Corvax International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PASTESEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PR1
; ORGANISM: Homo Sapien
US-09-776-191-2

Query Match 99.4%; Score 4653; DB 10; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKVNGLSEGVFPLPVNNVKKYKKGPGRWVLLAA 60
DB 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKVNGLSEGVFPLPVNNVKKYKKGPGRWVLLAA 60
QY 61 VLIGLLVLVIGLIGFLVWHLQYRDVAVQKXNGYMRITNENFVDAYENSNSSTEFVSLASKV 120
DB 61 VLIGLLVLVIGLIGFLVWHLQYRDVAVQKXNGYMRITNENFVDAYENSNSSTEFVSLASKV 120
QY 121 KDALKLISGVPLGPGYHKSAVTAPESEGVIAIYMWSEFSIPQHLVEARVMAEERVVM 180
DB 121 KDALKLISGVPLGPGYHKSAVTAPESEGVIAIYMWSEFSIPQHLVEARVMAEERVVM 180
QY 181 LPPPARSLKSFVVVSVAAPPTDSKTVORTODNSCSFGIHHARGVLMFTTGPFPDSYPYA 240
DB 181 LPPPARSLKSFVVVSVAAPPTDSKTVORTODNSCSFGIHHARGVLMFTTGPFPDSYPYA 240
QY 241 HARCQMLRGDADSVLSLTFERSFDLASCDERGSDLVTVYNTLSPMBEHALVOLCGTYPSP 300
DB 241 HARCQMLRGDADSVLSLTFERSFDLASCDERGSDLVTVYNTLSPMBEHALVOLCGTYPSP 300
QY 301 YNLTFHSSQNVLLITLTNTERRRHPGFPAFFPOLPRMSSCGRLRKAQGTNSPYFGHY 360
DB 301 YNLTFHSSQNVLLITLTNTERRRHPGFPAFFPOLPRMSSCGRLRKAQGTNSPYFGHY 360
```

```
QY 361 PENIDCTWNIIEVNNQHVKRFEFYLLIEPGVPAGTCEPDQYVEINKEKTCGERSQFVVTIS 420
DB 361 PENIDCTWNIIEVNNQHVKRFEFYLLIEPGVPAGTCEPDQYVEINKEKTCGERSQFVVTIS 420
QY 421 NSNKKITVRHSDOSYTDGTGLAEYLSYDSSDPCEQGFCTGTGACIKKEIRCDGMADCTTH 480
DB 421 NSNKKITVRHSDOSYTDGTGLAEYLSYDSSDPCEQGFCTGTGACIKKEIRCDGMADCTTH 480
QY 481 SDELNCSCDAGHOFCTCKNFKCKPLFWVCDSDVNDGNSDEQSCSPAOTFRCSNGKLSK 540
DB 481 SDELNCSCDAGHOFCTCKNFKCKPLFWVCDSDVNDGNSDEQSCSPAOTFRCSNGKLSK 540
QY 541 SQQCNKGDCCGSDSDEASCPKXNVVYCTKHTYRCINGLCLSKNPECDKEDCSDSDEK 600
DB 541 SQQCNKGDCCGSDSDEASCPKXNVVYCTKHTYRCINGLCLSKNPECDKEDCSDSDEK 600
QY 601 DCCGGLRSFTFRQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMWLVSAAHCYID 660
DB 601 DCCGGLRSFTFRQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMWLVSAAHCYID 660
QY 661 DRGFRYSDPQTQWTAFLGLHDQSORSAVQOERRLKRIISHPFNDFTFYDIALLEKXP 720
DB 661 DRGFRYSDPQTQWTAFLGLHDQSORSAVQOERRLKRIISHPFNDFTFYDIALLEKXP 720
QY 721 AEYSWVRPCLPDASHVPAGKAIWVTGHTQYGGTALLI0KGEIRVINQTCENLL 780
DB 721 AEYSWVRPCLPDASHVPAGKAIWVTGHTQYGGTALLI0KGEIRVINQTCENLL 780
QY 781 PQQITPRMVCVGLSGVDSCQDSGGLPSVEADGRIFGAGVVSMDGCAQRNKPQVYT 840
DB 781 PQQITPRMVCVGLSGVDSCQDSGGLPSVEADGRIFGAGVVSMDGCAQRNKPQVYT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

RESULT 9
US-10-099-700A-2
; Sequence 2, Application US/10099700A
; Publication No. US20030008372A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiumn-Chern Yeh
; APPLICANT: Corvax International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7, 1
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1613
; CURRENT APPLICATION NUMBER: US/10/099,700A
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/275,592
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PASTESEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PR1
; ORGANISM: Homo Sapien
US-10-099-700A-2

Query Match 99.4%; Score 4653; DB 14; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKVNGLSEGVFPLPVNNVKKYKKGPGRWVLLAA 60
DB 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKVNGLSEGVFPLPVNNVKKYKKGPGRWVLLAA 60
QY 61 VLIGLLVLVIGLIGFLVWHLQYRDVAVQKXNGYMRITNENFVDAYENSNSSTEFVSLASKV 120
DB 61 VLIGLLVLVIGLIGFLVWHLQYRDVAVQKXNGYMRITNENFVDAYENSNSSTEFVSLASKV 120
QY 121 KDALKLISGVPLGPGYHKSAVTAPESEGVIAIYMWSEFSIPQHLVEARVMAEERVVM 180
DB 121 KDALKLISGVPLGPGYHKSAVTAPESEGVIAIYMWSEFSIPQHLVEARVMAEERVVM 180
```

```
Db 121 KDALKLYSGVPFLGPHKESAVTAFASEGVIAYYMSFSLPQHLVEARVMAEERVVM 180
Qy 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSFGLHARGVELMFTTGPSPDPYPA 240
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSFGLHARGVELMFTTGPSPDPYPA 240
Qy 241 HARGOMLRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPMEPHALVOLCGTYPPS 300
Db 241 HARGOMLRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPMEPHALVOLCGTYPPS 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHHPGFEATFPOLPRMSSCGRLRKAQGTFSNPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHHPGFEATFPOLPRMSSCGRLRKAQGTFSNPPYPGHY 360
Qy 361 PPNIIDCTWNIIEVPNNQHVKVPFKPFYLLBPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Db 361 PPNIIDCTWNIIEVPNNQHVKVPFKPFYLLBPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Qy 421 NSNKITVRFHSDOSYTDGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDGMADCTDH 480
Db 421 NSNKITVRFHSDOSYTDGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDGMADCTDH 480
Qy 481 SDELINCSGDAGHQTCKNKKFCKPLFMVCDSVNDCGNSDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELINCSGDAGHQTCKNKKFCKPLFMVCDSVNDCGNSDEQSCSPAQTFRCNSGKCLSK 540
Qy 541 SQQNGKDDCGDGSDEASCPKVVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSEK 600
Db 541 SQQNGKDDCGDGSDEASCPKVVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSEK 600
Qy 601 DCDGGLSFTROARVVGSTDADEGEWPQVSLHALGGHICGASLISPNNLVSAHCTID 660
Db 601 DCDGGLSFTROARVVGSTDADEGEWPQVSLHALGGHICGASLISPNNLVSAHCTID 660
Qy 661 DRGFRYSDPQMTAFGLHDOSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEK 720
Db 661 DRGFRYSDPQMTAFGLHDOSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEK 720
Qy 721 AEYSMWVRPICLPDASHVFPAGKAIWVGWGHYGGTGALLQKGEIRVINOCTCENLL 780
Db 721 AEYSMWVRPICLPDASHVFPAGKAIWVGWGHYGGTGALLQKGEIRVINOCTCENLL 780
Qy 781 PQQITPRMVCVGLSGVDSCQSDSGPLSSVEADGRIFGAGVSVWGDGCAQRNKPQVYT 840
Db 781 PQQITPRMVCVGLSGVDSCQSDSGPLSSVEADGRIFGAGVSVWGDGCAQRNKPQVYT 840
Qy 841 RLPLFRDMIKENTGV 855
Db 841 RLPLFRDMIKENTGV 855

RESULT 10
US-10-190-030B-2
; Sequence 2, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
; FILE REFERENCE: 24745-1618
; CURRENT APPLICATION NUMBER: US/10/190, 030B
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-190-030B-2
```

```
Query Match 99.4%; Score 4653; DB 14; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MGSDBRARKGGGGKXDFGAGLKYNRSRHEKVNGLBEGVEFLPVNNVKYKXKGPGRMVLAA 60
Db 1 MGSDBRARKGGGGKXDFGAGLKYNRSRHEKVNGLBEGVEFLPVNNVKYKXKGPGRMVLAA 60
Qy 61 VLIGLLLVLLIGFLVNLHLDYRDYRVQKYNKGYRKTINENPVDAEENSNTFVSLSASKV 120
Db 61 VLIGLLLVLLIGFLVNLHLDYRDYRVQKYNKGYRKTINENPVDAEENSNTFVSLSASKV 120
Qy 121 KDALKLYSGVPFLGPHKESAVTAFASEGVIAYYMSFSLPQHLVEARVMAEERVVM 180
Db 121 KDALKLYSGVPFLGPHKESAVTAFASEGVIAYYMSFSLPQHLVEARVMAEERVVM 180
Qy 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSFGLHARGVELMFTTGPSPDPYPA 240
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSFGLHARGVELMFTTGPSPDPYPA 240
Qy 241 HARGOMLRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPMEPHALVOLCGTYPPS 300
Db 241 HARGOMLRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPMEPHALVOLCGTYPPS 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHHPGFEATFPOLPRMSSCGRLRKAQGTFSNPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHHPGFEATFPOLPRMSSCGRLRKAQGTFSNPPYPGHY 360
Qy 361 PPNIIDCTWNIIEVPNNQHVKVPFKPFYLLBPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Db 361 PPNIIDCTWNIIEVPNNQHVKVPFKPFYLLBPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Qy 421 NSNKITVRFHSDOSYTDGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDGMADCTDH 480
Db 421 NSNKITVRFHSDOSYTDGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDGMADCTDH 480
Qy 481 SDELINCSGDAGHQTCKNKKFCKPLFMVCDSVNDCGNSDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELINCSGDAGHQTCKNKKFCKPLFMVCDSVNDCGNSDEQSCSPAQTFRCNSGKCLSK 540
Qy 541 SQQNGKDDCGDGSDEASCPKVVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSEK 600
Db 541 SQQNGKDDCGDGSDEASCPKVVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSEK 600
Qy 601 DCDGGLSFTROARVVGSTDADEGEWPQVSLHALGGHICGASLISPNNLVSAHCTID 660
Db 601 DCDGGLSFTROARVVGSTDADEGEWPQVSLHALGGHICGASLISPNNLVSAHCTID 660
Qy 661 DRGFRYSDPQMTAFGLHDOSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEK 720
Db 661 DRGFRYSDPQMTAFGLHDOSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEK 720
Qy 721 AEYSMWVRPICLPDASHVFPAGKAIWVGWGHYGGTGALLQKGEIRVINOCTCENLL 780
Db 721 AEYSMWVRPICLPDASHVFPAGKAIWVGWGHYGGTGALLQKGEIRVINOCTCENLL 780
Qy 781 PQQITPRMVCVGLSGVDSCQSDSGPLSSVEADGRIFGAGVSVWGDGCAQRNKPQVYT 840
Db 781 PQQITPRMVCVGLSGVDSCQSDSGPLSSVEADGRIFGAGVSVWGDGCAQRNKPQVYT 840
Qy 841 RLPLFRDMIKENTGV 855
Db 841 RLPLFRDMIKENTGV 855

RESULT 11
US-10-302-840A-2
; Sequence 2, Application US/10302840A
; Publication No. US20030134794A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; APPLICANT: Ong, Edgar O.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCODED
```

```

; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1622
; CURRENT APPLICATION NUMBER: US/10/302,840A
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 60/332,015
; PRIOR FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-302-840A-2

Query Match      99.4%; Score 4653; DB 14; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 MGSBRARKGGGGKDFGAGLKYNSRHEKNGLEGEVEFLPVNNVKKYKKGPGRMVTLAA 60
DB      1 MGSBRARKGGGGKDFGAGLKYNSRHEKNGLEGEVEFLPVNNVKKYKKGPGRMVTLAA 60
QY      61 VLIGLLVLIGIFLVHMLQYRDVRYOKVKGVMRTINENFVDAYENSNSTEFVSLASKV 120
DB      61 VLIGLLVLIGIFLVHMLQYRDVRYOKVKGVMRTINENFVDAYENSNSTEFVSLASKV 120
QY      121 KDALKLLYSGVPLGPHKESAVTAFSEGSVIAYYSEFSIPQHLVEAEERVAEERVVM 180
DB      121 KDALKLLYSGVPLGPHKESAVTAFSEGSVIAYYSEFSIPQHLVEAEERVAEERVVM 180
QY      181 LPPRARSLSKSFVVTSSVAAPFTDSKTVOQTQDNCSEFGLHARGVELMRFTTGPFPDSYPYA 240
DB      181 LPPRARSLSKSFVVTSSVAAPFTDSKTVOQTQDNCSEFGLHARGVELMRFTTGPFPDSYPYA 240
QY      241 HARCOMALRGDADSVSLTFRSPDLASCDERGSIDLVTYVNTLSPMERHALVOLCGTTPPS 300
DB      241 HARCOMALRGDADSVSLTFRSPDLASCDERGSIDLVTYVNTLSPMERHALVOLCGTTPPS 300
QY      301 YNLTFFSSQNVLLITLTITNTERHHPGEATFFQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
DB      301 YNLTFFSSQNVLLITLTITNTERHHPGEATFFQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
QY      361 PPNIDCTWNIIEVNNQHVKKRFKFFYLLEPGVPAGTCPKDYVEINERKYCGERSQFVYVS 420
DB      361 PPNIDCTWNIIEVNNQHVKKRFKFFYLLEPGVPAGTCPKDYVEINERKYCGERSQFVYVS 420
QY      421 NSNKITVRFHSDOSYTDGTGLAEYLSYDSDPCPGQFTCRGTGIRKELRCDMACTDH 480
DB      421 NSNKITVRFHSDOSYTDGTGLAEYLSYDSDPCPGQFTCRGTGIRKELRCDMACTDH 480
QY      481 SDELINCSADAGHOFCTCKNFKCKPLFWVCDVNDGDNSEDGSCCPAQTFRCNGKCLSK 540
DB      481 SDELINCSADAGHOFCTCKNFKCKPLFWVCDVNDGDNSEDGSCCPAQTFRCNGKCLSK 540
QY      541 SQQNGNDGCDGSDASCPKVVNVVCTKTYRCLNGLCLSKNPEDECDGKEDCSGSDSK 600
DB      541 SQQNGNDGCDGSDASCPKVVNVVCTKTYRCLNGLCLSKNPEDECDGKEDCSGSDSK 600
QY      601 DCDGGLSFTROARVVGCTDADEGEWPMOYSLHALGQGHICGASLISPMNLVSAAHCIYD 660
DB      601 DCDGGLSFTROARVVGCTDADEGEWPMOYSLHALGQGHICGASLISPMNLVSAAHCIYD 660
QY      661 DRGRYSADPTQMTAFGLIHDSQSRAPGVDERRLKRIISHPFNFDTFYDIALLELEKP 720
DB      661 DRGRYSADPTQMTAFGLIHDSQSRAPGVDERRLKRIISHPFNFDTFYDIALLELEKP 720
QY      721 AEYSMSMRPITCLPASHVFPAGKAIWYTGWGHITQYGGTGALLQKGRIRYINQTTGNNL 780
DB      721 AEYSMSMRPITCLPASHVFPAGKAIWYTGWGHITQYGGTGALLQKGRIRYINQTTGNNL 780
QY      781 PQQITPRMMKCVGFLSGVDSCQDGGPGLSSVEADGRIFGAGVYVSWDGGCAGRRKPPVYT 840
DB      781 PQQITPRMMKCVGFLSGVDSCQDGGPGLSSVEADGRIFGAGVYVSWDGGCAGRRKPPVYT 840
```

```

QY      841 RLPLFRDMIKENTGV 855
DB      841 RLPLFRDMIKENTGV 855

RESULT 12
US-10-267-219-2
; Sequence 2, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ien, Jium-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2;
; FILE REFERENCE: 24745-1621
; CURRENT APPLICATION NUMBER: US/10/267,219
; CURRENT FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: 60/328,530
; PRIOR FILING DATE: 09-OCT-2001
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-267-219-2

Query Match      99.4%; Score 4653; DB 14; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 MGSBRARKGGGGKDFGAGLKYNSRHEKNGLEGEVEFLPVNNVKKYKKGPGRMVTLAA 60
DB      1 MGSBRARKGGGGKDFGAGLKYNSRHEKNGLEGEVEFLPVNNVKKYKKGPGRMVTLAA 60
QY      61 VLIGLLVLIGIFLVHMLQYRDVRYOKVKGVMRTINENFVDAYENSNSTEFVSLASKV 120
DB      61 VLIGLLVLIGIFLVHMLQYRDVRYOKVKGVMRTINENFVDAYENSNSTEFVSLASKV 120
QY      121 KDALKLLYSGVPLGPHKESAVTAFSEGSVIAYYSEFSIPQHLVEAEERVAEERVVM 180
DB      121 KDALKLLYSGVPLGPHKESAVTAFSEGSVIAYYSEFSIPQHLVEAEERVAEERVVM 180
QY      181 LPPRARSLSKSFVVTSSVAAPFTDSKTVOQTQDNCSEFGLHARGVELMRFTTGPFPDSYPYA 240
DB      181 LPPRARSLSKSFVVTSSVAAPFTDSKTVOQTQDNCSEFGLHARGVELMRFTTGPFPDSYPYA 240
QY      241 HARCOMALRGDADSVSLTFRSPDLASCDERGSIDLVTYVNTLSPMERHALVOLCGTTPPS 300
DB      241 HARCOMALRGDADSVSLTFRSPDLASCDERGSIDLVTYVNTLSPMERHALVOLCGTTPPS 300
QY      301 YNLTFFSSQNVLLITLTITNTERHHPGEATFFQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
DB      301 YNLTFFSSQNVLLITLTITNTERHHPGEATFFQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
QY      361 PPNIDCTWNIIEVNNQHVKKRFKFFYLLEPGVPAGTCPKDYVEINERKYCGERSQFVYVS 420
DB      361 PPNIDCTWNIIEVNNQHVKKRFKFFYLLEPGVPAGTCPKDYVEINERKYCGERSQFVYVS 420
QY      421 NSNKITVRFHSDOSYTDGTGLAEYLSYDSDPCPGQFTCRGTGIRKELRCDMACTDH 480
DB      421 NSNKITVRFHSDOSYTDGTGLAEYLSYDSDPCPGQFTCRGTGIRKELRCDMACTDH 480
QY      481 SDELINCSADAGHOFCTCKNFKCKPLFWVCDVNDGDNSEDGSCCPAQTFRCNGKCLSK 540
DB      481 SDELINCSADAGHOFCTCKNFKCKPLFWVCDVNDGDNSEDGSCCPAQTFRCNGKCLSK 540
QY      541 SQQNGNDGCDGSDASCPKVVNVVCTKTYRCLNGLCLSKNPEDECDGKEDCSGSDSK 600
DB      541 SQQNGNDGCDGSDASCPKVVNVVCTKTYRCLNGLCLSKNPEDECDGKEDCSGSDSK 600
QY      601 DCDGGLSFTROARVVGCTDADEGEWPMOYSLHALGQGHICGASLISPMNLVSAAHCIYD 660
```



Db 601 DCDGSLRFTROARVVGSTDADEGEMFWQVSLHALGGCHICGASLISPMNLVSAHCYID 660  
Qy 661 DRGRYSDPPTQMTAFGLHDQSORSAFGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720  
Db 661 DRGRYSDPPTQMTAFGLHDQSORSAFGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720  
Qy 721 AEYSMWRPICLPDASHVFPAGKAIWYTGWGHTOYGGTGALLQKGEIRVINOTTCEML 780  
Db 721 AEYSMWRPICLPDASHVFPAGKAIWYTGWGHTOYGGTGALLQKGEIRVINOTTCEML 780  
Qy 781 PQQITPRMCMVGFISGVDSQCGDSCGPLSSVEADGRIFGAGVSWGDGCAGRNPGVYT 840  
Db 781 PQQITPRMCMVGFISGVDSQCGDSCGPLSSVEADGRIFGAGVSWGDGCAGRNPGVYT 840  
Qy 841 RLPLFRDMIKENTGV 855  
Db 841 RLPLFRDMIKENTGV 855

RESULT 13  
US-10-112-221A-2  
; Sequence 2, Application US/10112221A  
; Publication No. US20030166851A1  
; GENERAL INFORMATION:  
; APPLICANT: Madison, Edwin  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9  
; FILE REFERENCE: 24745-1615  
; CURRENT APPLICATION NUMBER: US/10/112,221A  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/279,228  
; PRIOR FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/291,501  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-112-221A-2

Query Match 99.4%; Score 4653; DB 14; Length 855;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MGSRRARKGGGGPDPFAGLKYNSRHEKVNGLSEGVFLPVNNVKKVEKHGPGRWVTLAA 60  
Db 1 MGSRRARKGGGGPDPFAGLKYNSRHEKVNGLSEGVFLPVNNVKKVEKHGPGRWVTLAA 60  
Qy 61 VLIGLLVLLGIGFLVWHLQYRDVAVQKYNKGYMRITNENFVDAYENSNSTEFLVSLASKV 120  
Db 61 VLIGLLVLLGIGFLVWHLQYRDVAVQKYNKGYMRITNENFVDAYENSNSTEFLVSLASKV 120  
Qy 121 KDALKLTVSGVPFLGPRYHKEASAVTAFSEGSVIAYWSEFSIPQHLVEAEERVAEERVVM 180  
Db 121 KDALKLTVSGVPFLGPRYHKEASAVTAFSEGSVIAYWSEFSIPQHLVEAEERVAEERVVM 180  
Qy 181 LPPRARSLSKSFVAVSVAFPTDSKTVORTDONSQSGFGLHAGVVELMRPTTGPPDSPYPA 240  
Db 181 LPPRARSLSKSFVAVSVAFPTDSKTVORTDONSQSGFGLHAGVVELMRPTTGPPDSPYPA 240  
Qy 241 HARCOMALRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPMWPHALVOLCGTYPSS 300  
Db 241 HARCOMALRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPMWPHALVOLCGTYPSS 300  
Qy 301 YNLTFHSSQNVLLTLITNTERRHNGFEATFFQLPKMSCCGRRLKAKAGTNSPYYPGHY 360  
Db 301 YNLTFHSSQNVLLTLITNTERRHNGFEATFFQLPKMSCCGRRLKAKAGTNSPYYPGHY 360  
Qy 361 PPNIDCTNIEVPPNNQHVKVPFKFYLLBPGVPACTCPKDYVEINGEKYCGERSQFVVT 420

Db 361 PPNIDCTNIEVPPNNQHVKVPFKFYLLBPGVPACTCPKDYVEINGEKYCGERSQFVVT 420  
Qy 421 NSNKITTFRHSDDSYTTTGFLAEVLSYDSDPCPGQTCRTGRICRIGELRDCDGNADCTDH 480  
Db 421 NSNKITTFRHSDDSYTTTGFLAEVLSYDSDPCPGQTCRTGRICRIGELRDCDGNADCTDH 480  
Qy 481 SDELNGSCDGHQHTCKNKKCKPLFWVCDSVNDGSDNSDEQSCSPQOTFRCSNGKLSK 540  
Db 481 SDELNGSCDGHQHTCKNKKCKPLFWVCDSVNDGSDNSDEQSCSPQOTFRCSNGKLSK 540  
Qy 541 SQQNGKDDCGDSDASCPVNVVCTKITYRCLNGLCSKNGPECDDGKEDSGSDEK 600  
Db 541 SQQNGKDDCGDSDASCPVNVVCTKITYRCLNGLCSKNGPECDDGKEDSGSDEK 600  
Qy 601 DCDGSLRFTROARVVGSTDADEGEMFWQVSLHALGGCHICGASLISPMNLVSAHCYID 660  
Db 601 DCDGSLRFTROARVVGSTDADEGEMFWQVSLHALGGCHICGASLISPMNLVSAHCYID 660  
Qy 661 DRGRYSDPPTQMTAFGLHDQSORSAFGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720  
Db 661 DRGRYSDPPTQMTAFGLHDQSORSAFGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720  
Qy 721 AEYSMWRPICLPDASHVFPAGKAIWYTGWGHTOYGGTGALLQKGEIRVINOTTCEML 780  
Db 721 AEYSMWRPICLPDASHVFPAGKAIWYTGWGHTOYGGTGALLQKGEIRVINOTTCEML 780  
Qy 781 PQQITPRMCMVGFISGVDSQCGDSCGPLSSVEADGRIFGAGVSWGDGCAGRNPGVYT 840  
Db 781 PQQITPRMCMVGFISGVDSQCGDSCGPLSSVEADGRIFGAGVSWGDGCAGRNPGVYT 840  
Qy 841 RLPLFRDMIKENTGV 855  
Db 841 RLPLFRDMIKENTGV 855

RESULT 14  
US-10-104-271-2  
; Sequence 2, Application US/10104271  
; Publication No. US20030181658A1  
; GENERAL INFORMATION:  
; APPLICANT: Madison, Edwin  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP14, THE ENCODI  
; FILE REFERENCE: 24745-1614  
; CURRENT APPLICATION NUMBER: US/10/104,271  
; PRIOR FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/278,166  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-104-271-2

Query Match 99.4%; Score 4653; DB 14; Length 855;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MGSRRARKGGGGPDPFAGLKYNSRHEKVNGLSEGVFLPVNNVKKVEKHGPGRWVTLAA 60  
Db 1 MGSRRARKGGGGPDPFAGLKYNSRHEKVNGLSEGVFLPVNNVKKVEKHGPGRWVTLAA 60  
Qy 61 VLIGLLVLLGIGFLVWHLQYRDVAVQKYNKGYMRITNENFVDAYENSNSTEFLVSLASKV 120  
Db 61 VLIGLLVLLGIGFLVWHLQYRDVAVQKYNKGYMRITNENFVDAYENSNSTEFLVSLASKV 120  
Qy 121 KDALKLTVSGVPFLGPRYHKEASAVTAFSEGSVIAYWSEFSIPQHLVEAEERVAEERVVM 180  
Db 121 KDALKLTVSGVPFLGPRYHKEASAVTAFSEGSVIAYWSEFSIPQHLVEAEERVAEERVVM 180

```

QY 181 LPPRSLKSFVVTSVVAFPTDSKTQRTODNSCSFGIHAAGVELMFTTGPDPSPYA 240
|
|
|
Db 181 LPPRSLKSFVVTSVVAFPTDSKTQRTODNSCSFGIHAAGVELMFTTGPDPSPYA 240
QY 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSPLVTVYNTLSPMEPHALVOLCGTYPSP 300
|
|
|
Db 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSPLVTVYNTLSPMEPHALVOLCGTYPSP 300
QY 301 YNLTFRSSQVNLITLITNTERRHPPGEATFFQLPRMSCCGRLRKAQGTSPSPYRGHY 360
|
|
|
Db 301 YNLTFRSSQVNLITLITNTERRHPPGEATFFQLPRMSCCGRLRKAQGTSPSPYRGHY 360
QY 361 PPNIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420
|
|
|
Db 361 PPNIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420
QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSSDPGQFTCRGRCIRKELRCDGMADCTDH 480
|
|
|
Db 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSSDPGQFTCRGRCIRKELRCDGMADCTDH 480
QY 481 SDELNCSGADGHOFTCKNFKCKPLFWYCDSVNDCGNSDEQSCSPAQTFRCNSNGKCLSK 540
|
|
|
Db 481 SDELNCSGADGHOFTCKNFKCKPLFWYCDSVNDCGNSDEQSCSPAQTFRCNSNGKCLSK 540
QY 541 SQQCNKGKDCGSDSDASCPKRVNVVCTKATYRCLNGLCLSKNPECDGKEDCSGSDSK 600
|
|
|
Db 541 SQQCNKGKDCGSDSDASCPKRVNVVCTKATYRCLNGLCLSKNPECDGKEDCSGSDSK 600
QY 601 DCCGGLRSTFRQARVVGCTDADEGEWPWQVSLHALCGGHI CGASLISPNMLVSAAHCTYD 660
|
|
|
Db 601 DCCGGLRSTFRQARVVGCTDADEGEWPWQVSLHALCGGHI CGASLISPNMLVSAAHCTYD 660
QY 661 DRGFRYSPTQMTAFGLHDQSORSAPOVERLKRILISHPFNDFFDYDIALLELEXP 720
|
|
|
Db 661 DRGFRYSPTQMTAFGLHDQSORSAPOVERLKRILISHPFNDFFDYDIALLELEXP 720
QY 721 AEYSMWPRICLPDASHVFPAGKAIWVGWGHYQGTGALILQKEIRVINQTTCEML 780
|
|
|
Db 721 AEYSMWPRICLPDASHVFPAGKAIWVGWGHYQGTGALILQKEIRVINQTTCEML 780
QY 781 PQQITPRMVCVGLSGVDSQCGSGPLSVADGRIFGAGVSWGDGCAGRNKGVT 840
|
|
|
Db 781 PQQITPRMVCVGLSGVDSQCGSGPLSVADGRIFGAGVSWGDGCAGRNKGVT 840
QY 841 RLPLFRDMIKENTGV 855
|
|
|
Db 841 RLPLFRDMIKENTGV 855

RESULT 15
US-10-147-211A-2
; Sequence 2, Application US/10147211A
; Publication No. US20030235900A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edward
; APPLICANT: Yeh, Jiumn-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1
; FILE REFERENCE: 24745-1616
; CURRENT APPLICATION NUMBER: US/10/147,211A
; CURRENT FILING DATE: 2002-05-14
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-147-211A-2
Query Match 99.4%; Score 4653; DB 14; Length 855;

```

```

Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDBRARKGGGKPDGAGLKYNRSREKYNGLIEGVEFLPVNNVKKVEKHPRWVLLAA 60
|
|
|
Db 1 MGSDBRARKGGGKPDGAGLKYNRSREKYNGLIEGVEFLPVNNVKKVEKHPRWVLLAA 60
QY 61 VILGLLVLLIGFLVWMLQYRDVRYQKXNNGYRITNENFVADYENSSTEFVSLASKV 120
|
|
|
Db 61 VILGLLVLLIGFLVWMLQYRDVRYQKXNNGYRITNENFVADYENSSTEFVSLASKV 120
QY 121 KDLKLLYSQVPLGYPHKEASATLAFSEGVILAYVNSEPSIPQHLVEEARVMAEERVVM 180
|
|
|
Db 121 KDLKLLYSQVPLGYPHKEASATLAFSEGVILAYVNSEPSIPQHLVEEARVMAEERVVM 180
QY 181 LPPRSLKSFVVTSVVAFPTDSKTQRTODNSCSFGIHAAGVELMFTTGPDPSPYA 240
|
|
|
Db 181 LPPRSLKSFVVTSVVAFPTDSKTQRTODNSCSFGIHAAGVELMFTTGPDPSPYA 240
QY 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSPLVTVYNTLSPMEPHALVOLCGTYPSP 300
|
|
|
Db 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSPLVTVYNTLSPMEPHALVOLCGTYPSP 300
QY 301 YNLTFRSSQVNLITLITNTERRHPPGEATFFQLPRMSCCGRLRKAQGTSPSPYRGHY 360
|
|
|
Db 301 YNLTFRSSQVNLITLITNTERRHPPGEATFFQLPRMSCCGRLRKAQGTSPSPYRGHY 360
QY 361 PPNIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420
|
|
|
Db 361 PPNIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420
QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSSDPGQFTCRGRCIRKELRCDGMADCTDH 480
|
|
|
Db 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSSDPGQFTCRGRCIRKELRCDGMADCTDH 480
QY 481 SDELNCSGADGHOFTCKNFKCKPLFWYCDSVNDCGNSDEQSCSPAQTFRCNSNGKCLSK 540
|
|
|
Db 481 SDELNCSGADGHOFTCKNFKCKPLFWYCDSVNDCGNSDEQSCSPAQTFRCNSNGKCLSK 540
QY 541 SQQCNKGKDCGSDSDASCPKRVNVVCTKATYRCLNGLCLSKNPECDGKEDCSGSDSK 600
|
|
|
Db 541 SQQCNKGKDCGSDSDASCPKRVNVVCTKATYRCLNGLCLSKNPECDGKEDCSGSDSK 600
QY 601 DCCGGLRSTFRQARVVGCTDADEGEWPWQVSLHALCGGHI CGASLISPNMLVSAAHCTYD 660
|
|
|
Db 601 DCCGGLRSTFRQARVVGCTDADEGEWPWQVSLHALCGGHI CGASLISPNMLVSAAHCTYD 660
QY 661 DRGFRYSPTQMTAFGLHDQSORSAPOVERLKRILISHPFNDFFDYDIALLELEXP 720
|
|
|
Db 661 DRGFRYSPTQMTAFGLHDQSORSAPOVERLKRILISHPFNDFFDYDIALLELEXP 720
QY 721 AEYSMWPRICLPDASHVFPAGKAIWVGWGHYQGTGALILQKEIRVINQTTCEML 780
|
|
|
Db 721 AEYSMWPRICLPDASHVFPAGKAIWVGWGHYQGTGALILQKEIRVINQTTCEML 780
QY 781 PQQITPRMVCVGLSGVDSQCGSGPLSVADGRIFGAGVSWGDGCAGRNKGVT 840
|
|
|
Db 781 PQQITPRMVCVGLSGVDSQCGSGPLSVADGRIFGAGVSWGDGCAGRNKGVT 840
QY 841 RLPLFRDMIKENTGV 855
|
|
|
Db 841 RLPLFRDMIKENTGV 855

RESULT 16
US-10-156-214A-2
; Sequence 2, Application US/10156214A
; Publication No. US20040001801A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Joseph Edward Semple
; APPLICANT: George P. Vlasuk
; APPLICANT: Scott Jeffrey Kemp

```

```

; APPLICANT: Mallareddy Komandla
; APPLICANT: Daniel Vanna Siev
; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic Use
; TITLE OF INVENTION: Theeiof
; FILE REFERENCE: 24745-1611
; CURRENT APPLICATION NUMBER: US/10/156,214A
; CURRENT FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-156-214A-2

Query Match      99.4%; Score 4653; DB 15; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDBARKGGGKPKDFGAGLKYNSRHEKXNGLEGEVEFLPVNNVKKYKKGPGRWVTLAA 60
DB 1 MGSDBARKGGGKPKDFGAGLKYNSRHEKXNGLEGEVEFLPVNNVKKYKKGPGRWVTLAA 60
QY 61 VLIGLLVLLIGLGLVWHLQYRDYRVOKVXKXGYMRITNENFVDAEYNSNSTEFVSLASKV 120
DB 61 VLIGLLVLLIGLGLVWHLQYRDYRVOKVXKXGYMRITNENFVDAEYNSNSTEFVSLASKV 120
QY 121 KDALKLISGVPLGPIPHKSAVTAFSBGVIATYWSBPSIPQHLVBEAEVMAEEVVM 180
DB 121 KDALKLISGVPLGPIPHKSAVTAFSBGVIATYWSBPSIPQHLVBEAEVMAEEVVM 180
QY 121 KDALKLISGVPLGPIPHKSAVTAFSBGVIATYWSBPSIPQHLVBEAEVMAEEVVM 180
DB 121 KDALKLISGVPLGPIPHKSAVTAFSBGVIATYWSBPSIPQHLVBEAEVMAEEVVM 180
QY 181 LPPRARSLSFVWTSVVAAPFDSKTVOPTDONSFGHARGVGLMBPTTGPFDSYPYA 240
DB 181 LPPRARSLSFVWTSVVAAPFDSKTVOPTDONSFGHARGVGLMBPTTGPFDSYPYA 240
QY 181 LPPRARSLSFVWTSVVAAPFDSKTVOPTDONSFGHARGVGLMBPTTGPFDSYPYA 240
DB 181 LPPRARSLSFVWTSVVAAPFDSKTVOPTDONSFGHARGVGLMBPTTGPFDSYPYA 240
QY 241 HARCQALRGDADSVLSLTFRSPDLASCDERGSIDLVTYNTLSMEPHALVOLGTYPPS 300
DB 241 HARCQALRGDADSVLSLTFRSPDLASCDERGSIDLVTYNTLSMEPHALVOLGTYPPS 300
QY 241 HARCQALRGDADSVLSLTFRSPDLASCDERGSIDLVTYNTLSMEPHALVOLGTYPPS 300
DB 241 HARCQALRGDADSVLSLTFRSPDLASCDERGSIDLVTYNTLSMEPHALVOLGTYPPS 300
QY 301 YNLTFFSSQVNLTLITNTERRRHPGEATFPQLRPNSSCCGGRKAKAQGTNSYYGXY 360
DB 301 YNLTFFSSQVNLTLITNTERRRHPGEATFPQLRPNSSCCGGRKAKAQGTNSYYGXY 360
QY 301 YNLTFFSSQVNLTLITNTERRRHPGEATFPQLRPNSSCCGGRKAKAQGTNSYYGXY 360
DB 301 YNLTFFSSQVNLTLITNTERRRHPGEATFPQLRPNSSCCGGRKAKAQGTNSYYGXY 360
QY 361 PPNIDCWNTVEPNNOHVKKRPFKFFYLBPVPAGTCKOVVEINGEKYGESERGFVYTS 420
DB 361 PPNIDCWNTVEPNNOHVKKRPFKFFYLBPVPAGTCKOVVEINGEKYGESERGFVYTS 420
QY 361 PPNIDCWNTVEPNNOHVKKRPFKFFYLBPVPAGTCKOVVEINGEKYGESERGFVYTS 420
DB 361 PPNIDCWNTVEPNNOHVKKRPFKFFYLBPVPAGTCKOVVEINGEKYGESERGFVYTS 420
QY 421 NSNKITVRPHSDQSYTDGFLAEYLSYDSDPCGQFTCTGTCIRKELRCDGWADCTDH 480
DB 421 NSNKITVRPHSDQSYTDGFLAEYLSYDSDPCGQFTCTGTCIRKELRCDGWADCTDH 480
QY 421 NSNKITVRPHSDQSYTDGFLAEYLSYDSDPCGQFTCTGTCIRKELRCDGWADCTDH 480
DB 421 NSNKITVRPHSDQSYTDGFLAEYLSYDSDPCGQFTCTGTCIRKELRCDGWADCTDH 480
QY 481 SDELINSCDAGHQTCKXKFKPKLPWVCDVNDGSDSDEGSCCPAQTFCSNGKLSK 540
DB 481 SDELINSCDAGHQTCKXKFKPKLPWVCDVNDGSDSDEGSCCPAQTFCSNGKLSK 540
QY 481 SDELINSCDAGHQTCKXKFKPKLPWVCDVNDGSDSDEGSCCPAQTFCSNGKLSK 540
DB 481 SDELINSCDAGHQTCKXKFKPKLPWVCDVNDGSDSDEGSCCPAQTFCSNGKLSK 540
QY 541 SQQNGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDG 600
DB 541 SQQNGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDG 600
QY 541 SQQNGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDG 600
DB 541 SQQNGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDGSDG 600
QY 601 DCDGGLSFTROARVVGTDADDEGEMPOVSLHALGCGHLCGSLISPNMLVSAHICYID 660
DB 601 DCDGGLSFTROARVVGTDADDEGEMPOVSLHALGCGHLCGSLISPNMLVSAHICYID 660
QY 601 DCDGGLSFTROARVVGTDADDEGEMPOVSLHALGCGHLCGSLISPNMLVSAHICYID 660
DB 601 DCDGGLSFTROARVVGTDADDEGEMPOVSLHALGCGHLCGSLISPNMLVSAHICYID 660
QY 661 DRGRYSDDPTOWTAFLGLHDSQSRAPGVQERRLKRISHPFENDFTFYDIALLEKXP 720
DB 661 DRGRYSDDPTOWTAFLGLHDSQSRAPGVQERRLKRISHPFENDFTFYDIALLEKXP 720
QY 661 DRGRYSDDPTOWTAFLGLHDSQSRAPGVQERRLKRISHPFENDFTFYDIALLEKXP 720
DB 661 DRGRYSDDPTOWTAFLGLHDSQSRAPGVQERRLKRISHPFENDFTFYDIALLEKXP 720
QY 721 AYSVSWVRPCLPASHVPPAKAIWYTGWHTOYGGGALILQKGRVINTOTTENL 780
DB 721 AYSVSWVRPCLPASHVPPAKAIWYTGWHTOYGGGALILQKGRVINTOTTENL 780
QY 721 AYSVSWVRPCLPASHVPPAKAIWYTGWHTOYGGGALILQKGRVINTOTTENL 780
DB 721 AYSVSWVRPCLPASHVPPAKAIWYTGWHTOYGGGALILQKGRVINTOTTENL 780
QY 781 PQQITPRMVCVGLSGVDSCQDGSGLSVLEADGRIFGAGVSWDGCAGRNKPGVYT 840
DB 781 PQQITPRMVCVGLSGVDSCQDGSGLSVLEADGRIFGAGVSWDGCAGRNKPGVYT 840
```

```

DB 781 PQQITPRMVCVGLSGVDSCQDGSGLSVLEADGRIFGAGVSWDGCAGRNKPGVYT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

RESULT 17
US-10-072-012-352
; Sequence 352, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernov, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zethusen, Bryan
; APPLICANT: Ratturajan, Meera
; APPLICANT: Shukets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Beha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Futak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; REMAINING PRIOR Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 352
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-072-012-352

Query Match      99.4%; Score 4653; DB 15; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDBARKGGGKPKDFGAGLKYNSRHEKXNGLEGEVEFLPVNNVKKYKKGPGRWVTLAA 60
```

```

Db      1  MGSDRARKGGGGKDFAGLKYNSRHEKVNGLBEGVEFLPVNNVKYKKEKGPGRWVLLAA 60
Qy      61  VLLGLLVLLIGIFLVHMLQYRDVRYQKXNGYWRIRINENPVAYENSNGTEFVSLASKY 120
Db      61  VLLGLLVLLIGIFLVHMLQYRDVRYQKXNGYWRIRINENPVAYENSNGTEFVSLASKY 120
Qy      121  KDALKLLYSGVPLGPGYHKESAVTAFSEGSVIAIYWSFESIPOHLVEAERVAEERVVM 180
Db      121  KDALKLLYSGVPLGPGYHKESAVTAFSEGSVIAIYWSFESIPOHLVEAERVAEERVVM 180
Qy      181  LPPRARSLSKSFVVTSVVAFPTDSKTYQRTODNSCSFGLHARGVELMRFTTPGPPDSYPA 240
Db      181  LPPRARSLSKSFVVTSVVAFPTDSKTYQRTODNSCSFGLHARGVELMRFTTPGPPDSYPA 240
Qy      241  HARCQWALRGDADSVSLTFRSPDLASCDERGSDLVTYNTLSPMEPHALVOLCGTYPSP 300
Db      241  HARCQWALRGDADSVSLTFRSPDLASCDERGSDLVTYNTLSPMEPHALVOLCGTYPSP 300
Qy      301  YNLTFHSSQNVLLITLITNTERRHGPEATFPOLPRMSSCGRLRKAQGTENSPYYPGHY 360
Db      301  YNLTFHSSQNVLLITLITNTERRHGPEATFPOLPRMSSCGRLRKAQGTENSPYYPGHY 360
Qy      361  PPNIIDCTWNIENVNNOHVAKRFKFFYLLEBQVPACTPKQYVEINGRKXTCGESQFVYTS 420
Db      361  PPNIIDCTWNIENVNNOHVAKRFKFFYLLEBQVPACTPKQYVEINGRKXTCGESQFVYTS 420
Qy      421  NSNKITVRFHSDOSYDTGTGFLAEIYSVSDSPCGQGTCTGRCIRRELRCDDGADCTDH 480
Db      421  NSNKITVRFHSDOSYDTGTGFLAEIYSVSDSPCGQGTCTGRCIRRELRCDDGADCTDH 480
Qy      481  SDELINCSGADAGHOFCTCKNFKCKPLFWVCDVNDGDNDSDEQSCSPAQTFRCSSNGKCLSK 540
Db      481  SDELINCSGADAGHOFCTCKNFKCKPLFWVCDVNDGDNDSDEQSCSPAQTFRCSSNGKCLSK 540
Qy      541  SQCCNGDDDCGDSDEASCPKRVNVVCTKKTTCYCLNGLCLSKNPEDEGDESDGDEK 600
Db      541  SQCCNGDDDCGDSDEASCPKRVNVVCTKKTTCYCLNGLCLSKNPEDEGDESDGDEK 600
Qy      601  DCCGGLSFTROARVVGTDADBEEMPQVSLHALGCGHICGASLISPNMLVSAHACYID 660
Db      601  DCCGGLSFTROARVVGTDADBEEMPQVSLHALGCGHICGASLISPNMLVSAHACYID 660
Qy      661  DRGFRYSPTQWTAFLGLHDOSQASAPGVQERLRKRIISHPFNDFTFDYDIALLELEKP 720
Db      661  DRGFRYSPTQWTAFLGLHDOSQASAPGVQERLRKRIISHPFNDFTFDYDIALLELEKP 720
Qy      721  AEYSSWVRPCLDPAHSVFPAGKAIWTTGHTQYGGTGALLIQKGIIRIYNQTCENL 780
Db      721  AEYSSWVRPCLDPAHSVFPAGKAIWTTGHTQYGGTGALLIQKGIIRIYNQTCENL 780
Qy      781  PQGITPRMNCVGLSGVDSCQCGSGPLSSVEADGRIFGAGVYVSWGDGACAGNNKPPVYT 840
Db      781  PQGITPRMNCVGLSGVDSCQCGSGPLSSVEADGRIFGAGVYVSWGDGACAGNNKPPVYT 840
Qy      841  RLPLFRDWIKENTGV 855
Db      841  RLPLFRDWIKENTGV 855

```

```

; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier, Jr, Raymond J.
; APPLICANT: Gusev, Vladimil Y.
; APPLICANT: Wolenc, Steven D.
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Pena, Carol E. A
; APPLICANT: Grosbe, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072, 012
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265, 102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265, 514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265, 517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265, 412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265, 395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266, 406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266, 767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267, 057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266, 975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267, 459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 411
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-072-012-411

Query Match      99.4%; Score 4653; DB 15; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

RESULT 18
US-10-072-012-411
; Sequence 411, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shmukets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Esba
; APPLICANT: Padigattu, Muralidhara

```

```
DB 301 YNLTFFSSQNVLLITLITNTERRHPGEATFPQLPMSSCGRLRKAQGTFFNSPYGHY 360
QY 361 PENIDCTWNIIEVNNQHVKVRKFKFYLLBPGVAGTCKPDYVEINGEKYCGERSQFVVS 420
DB 361 PENIDCTWNIIEVNNQHVKVSFKFYLLBPGVAGTCKPDYVEINGEKYCGERSQFVVS 420
QY 421 NSNKITVRFHSDSYTDTGFLAEYLSYSSDPGQFTCTGRCIRKELACDGMADCTDH 480
DB 421 NSNKITVRFHSDSYTDTGFLAEYLSYSSDPGQFTCTGRCIRKELACDGMADCTDH 480
QY 481 SDELINSCDAGHOFCTCKNKECKPLFWYCDVNDGSDNDEOGSCCPAQTRCNSGKCLSK 540
DB 481 SDELINSCDAGHOFCTCKNKECKPLFWYCDVNDGSDNDEOGSCCPAQTRCNSGKCLSK 540
QY 541 SOQCNKGKDCGSDSDEASCPKRVNVVTCTKATYRCLNGLCLSKNPECDCGKEDSDGSD 600
DB 541 SOQCNKGKDCGSDSDEASCPKRVNVVTCTKATYRCLNGLCLSKNPECDCGKEDSDGSD 600
QY 601 DCCGGLRSFTRQARVVGTDADBEEMWQVSLHALGQGHICGASLISPNMLVSAHACYID 660
DB 601 DCCGGLRSFTRQARVVGTDADBEEMWQVSLHALGQGHICGASLISPNMLVSAHACYID 660
QY 661 DRGFRYSDPQMTAFGLGHDOSRABQVQERLKRILISHPFNDFTFDYDIALLELEXP 720
DB 661 DRGFRYSDPQMTAFGLGHDOSRABQVQERLKRILISHPFNDFTFDYDIALLELEXP 720
QY 721 AEYSSWVRPCLDPASHVFPAGKAIWVTGHTQYGGTGALLQKGEIRVYNQTTCEML 780
DB 721 AEYSSWVRPCLDPASHVFPAGKAIWVTGHTQYGGTGALLQKGEIRVYNQTTCEML 780
QY 781 PQOITPPMVCVGLSGVSDSCQDSSGRLSSVEADGILFGAGVYVWMDGCAQRKRPVYT 840
DB 781 PQOITPPMVCVGLSGVSDSCQDSSGRLSSVEADGILFGAGVYVWMDGCAQRKRPVYT 840
QY 841 RLPLFRDMIKENTGV 855
DB 841 RLPLFRDMIKENTGV 855

RESULT 19
; Sequence 418, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernyev, Velizar
; APPLICANT: Spytak, Kimberly
; APPLICANT: Zechusen, Bryan
; APPLICANT: Patcurajan, Meera
; APPLICANT: Shumkecs, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Legley, Denise M.
; APPLICANT: Rieget, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072, 012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265, 102
```

```
QY 1 MSDDRARKGGGKPKDAGLCKYNSRHEKYNGLSEGYEPLPVNNVKKYKGPBRWVLLA 60
DB 1 MSDDRARKGGGKPKDAGLCKYNSRHEKYNGLSEGYEPLPVNNVKKYKGPBRWVLLA 60
QY 61 VILGLLVLLGLGIFLVHMLQYRDVRYOKYNGYMRITNENFVDAYENSNSTEVSLSKY 120
DB 61 VILGLLVLLGLGIFLVHMLQYRDVRYOKYNGYMRITNENFVDAYENSNSTEVSLSKY 120
QY 121 KQALKLLYSVPLGPGYHKSATVAFSEGSVIAVYSEFISIPQHLVEAERVAEERVVM 180
DB 121 KQALKLLYSVPLGPGYHKSATVAFSEGSVIAVYSEFISIPQHLVEAERVAEERVVM 180
QY 181 LPPRASSLSFVTVSVAFPTDSKYQRTODNCSGFLHARGYELRFTTTPGPPDSPYPA 240
DB 181 LPPRASSLSFVTVSVAFPTDSKYQRTODNCSGFLHARGYELRFTTTPGPPDSPYPA 240
QY 241 HARCQWALRGDADSVSLTFRSFDLASCDERGSDLTVVNTTSLPMEPHALVOLCGTYPPS 300
DB 241 HARCQWALRGDADSVSLTFRSFDLASCDERGSDLTVVNTTSLPMEPHALVOLCGTYPPS 300
QY 301 YNLTFFSSQNVLLITLITNTERRHPGEATFPQLPMSSCGRLRKAQGTFFNSPYGHY 360
DB 301 YNLTFFSSQNVLLITLITNTERRHPGEATFPQLPMSSCGRLRKAQGTFFNSPYGHY 360
QY 361 PENIDCTWNIIEVNNQHVKVRKFKFYLLBPGVAGTCKPDYVEINGEKYCGERSQFVVS 420
DB 361 PENIDCTWNIIEVNNQHVKVSFKFYLLBPGVAGTCKPDYVEINGEKYCGERSQFVVS 420
QY 421 NSNKITVRFHSDSYTDTGFLAEYLSYSSDPGQFTCTGRCIRKELACDGMADCTDH 480
DB 421 NSNKITVRFHSDSYTDTGFLAEYLSYSSDPGQFTCTGRCIRKELACDGMADCTDH 480
QY 481 SDELINSCDAGHOFCTCKNKECKPLFWYCDVNDGSDNDEOGSCCPAQTRCNSGKCLSK 540
DB 481 SDELINSCDAGHOFCTCKNKECKPLFWYCDVNDGSDNDEOGSCCPAQTRCNSGKCLSK 540
QY 541 SOQCNKGKDCGSDSDEASCPKRVNVVTCTKATYRCLNGLCLSKNPECDCGKEDSDGSD 600
DB 541 SOQCNKGKDCGSDSDEASCPKRVNVVTCTKATYRCLNGLCLSKNPECDCGKEDSDGSD 600
QY 601 DCCGGLRSFTRQARVVGTDADBEEMWQVSLHALGQGHICGASLISPNMLVSAHACYID 660
DB 601 DCCGGLRSFTRQARVVGTDADBEEMWQVSLHALGQGHICGASLISPNMLVSAHACYID 660

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265, 514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265, 517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265, 412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265, 395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266, 406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266, 767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267, 057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266, 975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267, 459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 418
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-418

Query Match 99.4%; Score 4653; DB 15; Length 855;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

Db 601 DCCGSRFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLVSAAHCYID 660  
Qy 661 DRGFRYSDEPTQWTAFLGLHDQSORSAPOVERLKRIIISHPFNDFTFYDIALLELEXP 720  
Db 661 DRGFRYSDEPTQWTAFLGLHDQSORSAPOVERLKRIIISHPFNDFTFYDIALLELEXP 720  
Qy 721 AEYSMWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALLIOKGEIRVINOCTCENLL 780  
Db 721 AEYSMWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALLIOKGEIRVINOCTCENLL 780  
Qy 781 PQQITPRMNCVGLSGGVDSQQDSSGGLSSVEADGRIFGAGVSWGDCGAGRNPQVYT 840  
Db 781 PQQITPRMNCVGLSGGVDSQQDSSGGLSSVEADGRIFGAGVSWGDCGAGRNPQVYT 840  
Qy 841 RLPLFRDWIKENTGV 855  
Db 841 RLPLFRDWIKENTGV 855

## RESULT 20

US-10-600-187-2  
; Sequence 2, Application US/10600187  
; Publication No. US20040086910A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Tanimoto, Hirotsoshi  
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
; TITLE OF INVENTION: Overexpressed in Caichinomas  
; FILE REFERENCE: D6064CIF/D  
; CURRENT APPLICATION NUMBER: US/10/600,187  
; CURRENT FILING DATE: 2003-06-20  
; PRIOR APPLICATION NUMBER: US/09/654,600A  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 09/421,213  
; 09/027,337  
; PRIOR FILING DATE: 1999-10-20  
; 1998-02-20  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: TADG-15  
US-10-600-187-2

Query Match 99.4%; Score 4653; DB 15; Length 855;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MGSDDRARKGGGKDFGAGLKXNSRHEKXVGLGEGVEFLPVNNVKKVEKHGPGRWVTLAA 60  
Db 1 MGSDDRARKGGGKDFGAGLKXNSRHEKXVGLGEGVEFLPVNNVKKVEKHGPGRWVTLAA 60  
Qy 61 VLLGLLVLLIGIFLVWHLQYRDVAVQKXNGYMRITNENFVDAEYNSNSTEFVSLASKV 120  
Db 61 VLLGLLVLLIGIFLVWHLQYRDVAVQKXNGYMRITNENFVDAEYNSNSTEFVSLASKV 120  
Qy 121 KDLAKLLYSVPLFGPHKXSANTAFSEGSVIAYMSEFSIPDHLVEAEERVAEERVVM 180  
Db 121 KDLAKLLYSVPLFGPHKXSANTAFSEGSVIAYMSEFSIPDHLVEAEERVAEERVVM 180  
Qy 181 LPPRASLSKSFVYTSVVAAPTDSTKVTQRTQDNSCFGLHARGVLEMFTRPGPDSPPYA 240  
Db 181 LPPRASLSKSFVYTSVVAAPTDSTKVTQRTQDNSCFGLHARGVLEMFTRPGPDSPPYA 240  
Qy 241 HARCQWALRGDADSVLSLTFRSFDLASCDESGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300  
Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDESGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300  
Qy 301 YNLTFFSQVALLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTFFNSPYYPGHY 360  
Db 301 YNLTFFSQVALLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTFFNSPYYPGHY 360

Db 301 YNLTFFSQVALLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTFFNSPYYPGHY 360  
Qy 361 PNVIDCTWNIIEVNNOHVYRPFKFFYLLEPGVAGTCCPDQVYEINGEKYCGERSQFVYTS 420  
Db 361 PNVIDCTWNIIEVNNOHVYRPFKFFYLLEPGVAGTCCPDQVYEINGEKYCGERSQFVYTS 420  
Qy 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGTGCIKRELRCGDMADCTDH 480  
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGTGCIKRELRCGDMADCTDH 480  
Qy 481 SDFLNSCDAGHQFTCKNFKCPLFWVCDSDVNDCCGNSDEQSCGCAQTFRCNGKCLSK 540  
Db 481 SDFLNSCDAGHQFTCKNFKCPLFWVCDSDVNDCCGNSDEQSCGCAQTFRCNGKCLSK 540  
Qy 541 SQOCNGKDDCGDSDSASCPKXNVVVTCTKHTRCLNGLCLSKGNPCDQKEDCSDSDEK 600  
Db 541 SQOCNGKDDCGDSDSASCPKXNVVVTCTKHTRCLNGLCLSKGNPCDQKEDCSDSDEK 600  
Qy 601 DCDGSRFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLVSAAHCYID 660  
Db 601 DCDGSRFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLVSAAHCYID 660  
Qy 661 DRGFRYSDEPTQWTAFLGLHDQSORSAPOVERLKRIIISHPFNDFTFYDIALLELEXP 720  
Db 661 DRGFRYSDEPTQWTAFLGLHDQSORSAPOVERLKRIIISHPFNDFTFYDIALLELEXP 720  
Qy 721 AEYSMWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALLIOKGEIRVINOCTCENLL 780  
Db 721 AEYSMWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALLIOKGEIRVINOCTCENLL 780  
Qy 781 PQQITPRMNCVGLSGGVDSQQDSSGGLSSVEADGRIFGAGVSWGDCGAGRNPQVYT 840  
Db 781 PQQITPRMNCVGLSGGVDSQQDSSGGLSSVEADGRIFGAGVSWGDCGAGRNPQVYT 840  
Qy 841 RLPLFRDWIKENTGV 855  
Db 841 RLPLFRDWIKENTGV 855

## RESULT 21

US-10-276-774-1798  
; Sequence 1798, Application US/10276774  
; Publication No. US20040053245A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; APPLICANT: Tang, Y. Tom et al  
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-030  
; CURRENT APPLICATION NUMBER: US/10/276,774  
; CURRENT FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: 09/560,875  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; NUMBER OF SEQ ID NOS: 2700  
; SOFTWARE: Custom  
; SEQ ID NO 1798  
; LENGTH: 851  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-276-774-1798

Query Match 98.5%; Score 4614; DB 15; Length 851;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 843; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 7 RKGGGKDFGAGLKXNSRHEKXVGLGEGVEFLPVNNVKKVEKHGPGRWVTLAAVILIGLL 66  
Db 3 RKGGGKDFGAGLKXNSRHEKXVGLGEGVEFLPVNNVKKVEKHGPGRWVTLAAVILIGLL 62  
Qy 67 LVLLIGIFLVWHLQYRDVAVQKXNGYMRITNENFVDAEYNSNSTEFVSLASKVQDALKL 126  
Db 67 LVLLIGIFLVWHLQYRDVAVQKXNGYMRITNENFVDAEYNSNSTEFVSLASKVQDALKL 122

127 LYGVPFLGPHYKESAVTAESGVIAAYMSEPSIPOHLYEERVAEERVMLEPPRAR 186  
123 LYGVPFLGPHYKESAVTAESGVIAAYMSEPSIPOHLYEERVAEERVMLEPPRAR 182  
187 SLKSFVVTSVVAFPTDSKTQVQRTQDSCSFGLHARGVELMRFTTPGFPDSPYPAHARQW 246  
183 SLKSFVVTSVVAFPTDSKTQVQRTQDSCSFGLHARGVELMRFTTPGFPDSPYPAHARQW 242  
247 ALRGDADSVLSLTFRSFDLASCDERGRHLVTVYNTLSPMEPHALVOLCGTYPPSYNLTFFH 306  
243 ALRGDADSVLSLTFRSFDLASCDERGRHLVTVYNTLSPMEPHALVOLCGTYPPSYNLTFFH 302  
307 SSONVLLITLITINTERHHPGEATFFQLPNMSCGRLRAOQTFNSPYRGHYPPNIDC 366  
303 SSONVLLITLITINTERHHPGEATFFQLPNMSCGRLRAOQTFNSPYRGHYPPNIDC 362  
367 TNNIEVNNQHVKKRFFKFFYLEPGVPAGTCCPDYVEINKEKYGERSQFVYTSNSNKIT 426  
363 TNNIEVNNQHVKKRFFKFFYLEPGVPAGTCCPDYVEINKEKYGERSQFVYTSNSNKIT 422  
427 VRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGTGRCIKELRCGMDCTDHSDELNC 486  
423 VRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGTGRCIKELRCGMDCTDHSDELNC 482  
487 SCAGHQFTCKNFKCKPLFWVCDLVNDGDNSEOGSCPAQTFRCNGKCLSKSQCCNG 546  
483 SCAGHQFTCKNFKCKPLFWVCDLVNDGDNSEOGSCPAQTFRCNGKCLSKSQCCNG 542  
547 KDDCGSDDEASCPKNNVVTCTKHTYRCLNGLCLSKGNPECDGKEDSDSDKDCDGL 606  
543 KDDCGSDDEASCPKNNVVTCTKHTYRCLNGLCLSKGNPECDGKEDSDSDKDCDGL 602  
607 RSFTROARVVGGTDADEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDRGFRY 666  
603 RSFTROARVVGGTDADEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDRGFRY 662  
667 SDPTQWTAFLGLHDQSORSAFGVQERLKRILSHPFNDFTFYDIALLELEKPAEYSSM 726  
663 SDPTQWTAFLGLHDQSORSAFGVQERLKRILSHPFNDFTFYDIALLELEKPAEYSSM 722  
727 VRPCLPDASHVPAGKAIWVTGNGHTQYGGTALILQKGEIRVINOCTCENLLPQOITP 786  
723 VRPCLPDASHVPAGKAIWVTGNGHTQYGGTALILQKGEIRVINOCTCENLLPQOITP 782  
787 RMMCVGFLSGGVDSGCGDSSGGLSVADGRIFGAGVSVWGDCAGRNKPGVYTRLPFLR 846  
783 RMMCVGFLSGGVDSGCGDSSGGLSVADGRIFGAGVSVWGDCAGRNKPGVYTRLPFLR 842  
847 DWIKENTGV 855  
843 DWIKENTGV 851

RESULT 22  
US-10-296-115-1143  
; Sequence 1143, Application US/10296115  
; Publication No. US20040053248A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq Inc  
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides  
; FILE REFERENCE: 784PCT  
; CURRENT APPLICATION NUMBER: US/10/296,115  
; PRIOR APPLICATION NUMBER: 2002-11-18  
; PRIOR FILING DATE: 2000-01-21  
; PRIOR APPLICATION NUMBER: US09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 1478  
; SEQ ID NO 1143  
; LENGTH: 851  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-296-115-1143

Query Match 98.5%; Score 4614; DB 15; Length 851;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 843; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

7 RKGGGPKDGAALKTNSREHEKYNGLSEGEFLPVANNVKKVEKGGRWVLAAILGLL 66  
3 RKGGGPKDGAALKTNSREHEKYNGLSEGEFLPVANNVKKVEKGGRWVLAAILGLL 62  
67 LVLLIGFLVMHLQYDVRVQKVRNGYMRITNENPVDAEYNSSTFVSLASVKDALKL 126  
63 LVLLIGFLVMHLQYDVRVQKVRNGYMRITNENPVDAEYNSSTFVSLASVKDALKL 122  
127 LYGVPFLGPHYKESAVTAESGVIAAYMSEPSIPOHLYEERVAEERVMLEPPRAR 186  
123 LYGVPFLGPHYKESAVTAESGVIAAYMSEPSIPOHLYEERVAEERVMLEPPRAR 182  
187 SLKSFVVTSVVAFPTDSKTQVQRTQDSCSFGLHARGVELMRFTTPGFPDSPYPAHARQW 246  
183 SLKSFVVTSVVAFPTDSKTQVQRTQDSCSFGLHARGVELMRFTTPGFPDSPYPAHARQW 242  
247 ALRGDADSVLSLTFRSFDLASCDERGRHLVTVYNTLSPMEPHALVOLCGTYPPSYNLTFFH 306  
243 ALRGDADSVLSLTFRSFDLASCDERGRHLVTVYNTLSPMEPHALVOLCGTYPPSYNLTFFH 302  
307 SSONVLLITLITINTERHHPGEATFFQLPNMSCGRLRAOQTFNSPYRGHYPPNIDC 366  
303 SSONVLLITLITINTERHHPGEATFFQLPNMSCGRLRAOQTFNSPYRGHYPPNIDC 362  
367 TNNIEVNNQHVKKRFFKFFYLEPGVPAGTCCPDYVEINKEKYGERSQFVYTSNSNKIT 426  
363 TNNIEVNNQHVKKRFFKFFYLEPGVPAGTCCPDYVEINKEKYGERSQFVYTSNSNKIT 422  
427 VRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGTGRCIKELRCGMDCTDHSDELNC 486  
423 VRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGTGRCIKELRCGMDCTDHSDELNC 482  
487 SCAGHQFTCKNFKCKPLFWVCDLVNDGDNSEOGSCPAQTFRCNGKCLSKSQCCNG 546  
483 SCAGHQFTCKNFKCKPLFWVCDLVNDGDNSEOGSCPAQTFRCNGKCLSKSQCCNG 542  
547 KDDCGSDDEASCPKNNVVTCTKHTYRCLNGLCLSKGNPECDGKEDSDSDKDCDGL 606  
543 KDDCGSDDEASCPKNNVVTCTKHTYRCLNGLCLSKGNPECDGKEDSDSDKDCDGL 602  
607 RSFTROARVVGGTDADEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDRGFRY 666  
603 RSFTROARVVGGTDADEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDRGFRY 662  
667 SDPTQWTAFLGLHDQSORSAFGVQERLKRILSHPFNDFTFYDIALLELEKPAEYSSM 726  
663 SDPTQWTAFLGLHDQSORSAFGVQERLKRILSHPFNDFTFYDIALLELEKPAEYSSM 722  
727 VRPCLPDASHVPAGKAIWVTGNGHTQYGGTALILQKGEIRVINOCTCENLLPQOITP 786  
723 VRPCLPDASHVPAGKAIWVTGNGHTQYGGTALILQKGEIRVINOCTCENLLPQOITP 782  
787 RMMCVGFLSGGVDSGCGDSSGGLSVADGRIFGAGVSVWGDCAGRNKPGVYTRLPFLR 846  
783 RMMCVGFLSGGVDSGCGDSSGGLSVADGRIFGAGVSVWGDCAGRNKPGVYTRLPFLR 842  
847 DWIKENTGV 855  
843 DWIKENTGV 851

RESULT 23  
US-10-729-807-1  
; Sequence 1, Application US/10729807  
; Publication No. US20040132158A1  
; GENERAL INFORMATION:  
; APPLICANT: BANDMAN, Olga; HILLMAN, Jennifer L.



```
APPLICANT: TANG, Y. Tom, IAL, Preeti G.
APPLICANT: YUE, Henry; AZIMZAI, Yalda
APPLICANT: BAUGHN, Mariah R.; LU, Dyung Aina M.
TITLE OF INVENTION: HUMAN PEPTIDASES
FILE REFERENCE: PR-0651-1 DIV
CURRENT APPLICATION NUMBER: US/10/729, 807
CURRENT FILING DATE: 2003-12-05
PRIOR APPLICATION NUMBER: US 09/889,238
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: PCT/US00/00641
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: US 60/172,247
PRIOR FILING DATE: 1999-01-11
PRIOR APPLICATION NUMBER: US 60/132,253
PRIOR FILING DATE: 1999-05-03
PRIOR APPLICATION NUMBER: US 60/136,653
PRIOR FILING DATE: 1999-05-27
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 762
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No: 155179CD1
US-10-729-807-1
```

Query Match 89.0%; Score 4167; DB 16; Length 762;

Best Local Similarity 99.6%; Pred. No. 3,6e-310;

Matches 759; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
94 MRITNFVDAYENSSTEFVSLASKYKDALKLLYSGVPLGPHKESAVTASEGSAVIA 153
1 MRTINEFVAYENSSTEFVSLASKYKDALKLLYSVPLGPHKESAVTASEGSAVIA 60
154 YWSEFSIPOHLVEAEARVNAEERVVMPPRARSLSKSFVTVSVAFPTDSKTVQRTDMS 213
61 YWSEFSIPOHLVEAEARVNAEERVVMPPRARSLSKSFVTVSVAFPTDSKTVQRTDMS 120
214 CSFGIARAGVEMLRFTTGGFPDSDYPNARCOMALRDADSVLSLTFRSFDLASCDGRGS 273
121 CSFGIARAGVEMLRFTTGGFPDSDYPNARCOMALRDADSVLSLTFRSFDLASCDGRGS 180
274 DLVTYVNTLSPEMEPHALVOLCGTYPSPSYNTLFHSSQVNLITLITNTERHHPGEATFPQ 333
181 DLVTYVNTLSPEMEPHALVOLCGTYPSPSYNTLFHSSQVNLITLITNTERHHPGEATFPQ 240
334 LPRMSSCGRLRKAQGTFSNPPYRGHYPNIDCTWNIIEVNNQHVKRFKFFYLLBGPV 393
241 LPRMSSCGRLRKAQGTFSNPPYRGHYPNIDCTWNIIEVNNQHVKRFKFFYLLBGPV 300
394 AGTCPPDYVINEKYGERSQFVNTSNKITYRHFHSDSYNDTGLAFLYSDSDPC 453
301 AGTCPPDYVINEKYGERSQFVNTSNKITYRHFHSDSYNDTGLAFLYSDSDPC 360
454 PGOFTCTGACIRKEALRCDMADCTDHSDELNCSCDAGHOFCTCKNKCCKPLFWCDSVND 513
361 PGOFTCTGACIRKEALRCDMADCTDHSDELNCSCDAGHOFCTCKNKCCKPLFWCDSVND 420
514 CGNSDDEQSCSPAQFTRCSNGKCLSKSQCNKGKDCGSDSDASCPKRVNVVTCTKTYR 573
421 CGNSDDEQSCSPAQFTRCSNGKCLSKSQCNKGKDCGSDSDASCPKRVNVVTCTKTYR 480
574 CLNGCLCSKNPDEDDGSDSDSDGDEKDCGGLASFTROARVVGCTPADGEMPMQVSLH 633
481 CLNGCLCSKNPDEDDGSDSDSDGDEKDCGGLASFTROARVVGCTPADGEMPMQVSLH 540
634 ALQCGHICGASLISPNLVSAAHICYIDRGFRYSDDPTOMTAFGLHDOSORSAAGVOERR 633
541 ALQCGHICGASLISPNLVSAAHICYIDRGFRYSDDPTOMTAFGLHDOSORSAAGVOERR 600
694 LKRIISHPFNDFTFDYDIALLELEKPAEYSSMWVRICLPDASHVFPAGKAIWVTGMGHT 753
```

```
|||||
601 LKRIISHPFNDFTFDYDIALLELEKPAEYSSMWVRICLPDASHVFPAGKAIWVTGMGHT 660
754 QVCGTALLIQGEIRVINOTTCENLLPOIIPRMMCVGFLSGVNSCGDSCGFLPSVYE 813
661 QVCGTALLIQGEIRVINOTTCENLLPOIIPRMMCVGFLSGVNSCGDSCGFLPSVYE 720
814 ADGRIFAGVSWGDCAGRNKRGVYTRLPFLRDMIKENTGV 855
721 ADGRIFAGVSWGDCAGRNKRGVYTRLPFLRDMIKENTGV 762
```

RESULT 24

US-10-097-340-312

Sequence 312, Application US/10097340

Publication No. US20030087250A1

GENERAL INFORMATION:

APPLICANT: JOHN MONAHAN

APPLICANT: Manjula GANNAVAPU

APPLICANT: Sebastian HOERSCH

APPLICANT: Shubhangi KAMATKAR

APPLICANT: Steve G. KOVATS

APPLICANT: Rachel E. MEYERS

APPLICANT: Michael MORRISBY

APPLICANT: Peter OLANDT

APPLICANT: Aml SEN

APPLICANT: Peter VERBY

APPLICANT: Gordon B. MILLS

APPLICANT: Robert C. BAST, Jr.

APPLICANT: Karen LU

APPLICANT: Rosemarie SCHMANDT

APPLICANT: Xumei ZHAO

APPLICANT: Karen GLATT

TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification, Assessment, Prevention, and Therapy of Ovarian Cancer

FILE REFERENCE: MRI-030

CURRENT APPLICATION NUMBER: US/10/097,340

CURRENT FILING DATE: 2002-03-14

PRIOR APPLICATION NUMBER: 60/276, 025

PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: 60/325,149

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: 60/276, 026

PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: 60/324, 967

PRIOR FILING DATE: 2001/09/26

PRIOR APPLICATION NUMBER: 60/311, 732

PRIOR FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: 60/325,102

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: 60/323, 580

PRIOR FILING DATE: 2001-09-19

NUMBER OF SEQ ID NOS: 363

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 312

LENGTH: 782

TYPE: PRT

ORGANISM: Homo sapiens

US-10-097-340-312

Query Match 88.8%; Score 4158.5; DB 14; Length 782;

Best Local Similarity 90.9%; Pred. No. 1.7e-309;

Matches 777; Conservative 1; Mismatches 4; Indels 73; Gaps 2;

```
1 MGSDBARKKGGGKPKDGAAGLKNYSREKNGLEEGVEFLPVNNVKKVKGPGRWVLLA 60
1 MGSDBARKKGGGKPKDGAAGLKNYSREKNGLEEGVEFLPVNNVKKVKGPGRWVLLA 60
61 VILGLLVLLIGLGFVLMHTQYRDVRYQKYNQGYMRITNENFVDAYENSSTEFVSLASKY 120
61 VILGLLVLLIGLGFVLMHTQYRDVRYQKYNQGYMRITNENFVDAYENSSTEFVSLASKY 120
121 KDALKLLYSGVPLGPHKESAVTASEGSAVIAYWSEFSIPOHLVEAEARVNAEERVVM 180
```

```

Db      121 KDAKLKLYSGVPLGPHKESAVTAPEGSVIATYMSFSI:POLVLEAEAEVMAEERVVM 180
Qy      181 LPPARSLKSFVYTSVVAAPFTDSTKVORTQDNCSFGI:HAAGVELMRAFTTGPDPSPYPA 240
Db      181 LPPARSLKSFVYTSVVAAPFTDSTKVORTQDNCSFGI:HAAGVELMRAFTTGPDPSPYPA 240
Qy      241 HARCQMLRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLS:PMERHALVOLCGTYPSP 300
Db      241 HARCQMLRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLS:PMERHALVOLCGTYPSP 300
Qy      301 YNLTFSSQVLLITLITNTERRHGFEATFQQLPRMS:CGRLRKAQGTFSPPYPGHY 360
Db      301 YNLTFSSQVLLITLITNTERRHGFEATFQQLPRMS:CGRLRKAQGTFSPPYPGHY 360
Qy      361 PENIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPKDYVE:INGEKYCGERSQFVYTS 420
Db      361 PENIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPKDYVE:INGEKYCGERSQFVYTS 420
Qy      421 NSNKTIVRFHSDSYTDGTGFLAEVLSYDSSDPCPGQFTCR:GRCIRKELRCDMADCTDH 480
Db      421 NSNKTIVRFHSDSYTDGTGFLAEVLSYDSSDPCPGQFTCR:GRCIRKELRCDMADCTDH 480
Qy      481 SDELNCSDAHQHTCKNKKFCKPLFWVCDSVNDCGNSD:EQSCSPAQTFRCNSGKCLSK 540
Db      481 SDELNCSDAHQHTCKNKKFCKPLFWVCDSVNDCGNSD:EQSCSPAQTFRCNSGKCLSK 540
Qy      541 SQQCGKDDCGSDSDASCPANVVTCTKATYRCLNGI:CLSKNPECDKEDKSDGSD 600
Db      541 SQQCGKDDCGSDSDASCPANVVTCTKATYRCLNGI:CLSKNPECDKEDKSDGSD 600
Qy      601 DCCGSLFTRQARVNGTADDEGEMPVQVSLHAGCH:CGASLISPMNLVSAHCYID 660
Db      601 DCCGSLFTRQARVNGTADDEGEMPVQVSLHAGCH:CGASLISPMNLVSAHCYID 660
Qy      661 DRGRYSADPTQWTAFLGLHDQSRAPGVQERRLKRII:SHPFNDFTFDYDIALLEKRP 720
Db      661 DRGRYSADPTQWTAFLGLHDQSRAPGVQERRLKRII:SHPFNDFTFDYDIALLEKRP 720
Qy      721 AEYSWVRPCLPDASHVFPAGKAIWYTGWHTGYGGT:GALLIOKGEIRVINTQTCENL 780
Db      721 AEYSWVRPCLPDASHVFPAGKAIWYTGWHTGYGGT:GALLIOKGEIRVINTQTCENL 780
Qy      781 PQQITPRMCCVFLSGVDSQSGSPLSVLEADGRI:FGAGVYSWQDGCGRKRPVYT 840
Db      781 PQQITPRMCCVFLSGVDSQSGSPLSVLEADGRI:FGAGVYSWQDGCGRKRPVYT 840
Qy      841 RLPLFRDMIKENTGV 855
Db      841 RLPLFRDMIKENTGV 855
Qy      768 RLPLFRDMIKENTGV 782
Db      768 RLPLFRDMIKENTGV 782

```

```

APPLICANT: Pena, Carol E. A
APPLICANT: Futrak, Katarzyna
APPLICANT: Groose, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OR INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 44
LENGTH: 757
TYPE: PRT
ORGANISM: Homo sapiens
US-10-072-012-44

Query Match      87.4%; Score 4092; DB 15; Length 757;
Best Local Similarity 88.2%; Pred. No. 2e-304;
Matches 754; Conservative 0; Mismatches 3; Indels 98; Gaps 1;

Qy      1  MGSDRARKGGGGRPRDAGLAKYNSRHKVNGLEBGVFLPVNNVKKYKKGKPGRRVYLLA 60
Db      1  MGSDRARKGGGGRPRDAGLAKYNSRHKVNGLEBGVFLPVNNVKKYKKGKPGRRVYLLA 60
Qy      61  VLIIGLLVLIIGIGFLVWHLQYRUVRVQKVGKGYMRITNEFVDAYENSNTBFVSLASKV 120
Db      61  VLIIGLLVLIIGIGFLVWHLQYRUVRVQKVGKGYMRITNEFVDAYENSNTBFVSLASKV 120
Qy      61  VLIIGLLVLIIGIGFLVWHLQYRUVRVQKVGKGYMRITNEFVDAYENSNTBFVSLASKV 120
Db      61  VLIIGLLVLIIGIGFLVWHLQYRUVRVQKVGKGYMRITNEFVDAYENSNTBFVSLASKV 120
Qy      121 KDAKLKLYSGVPLGPHKESAVTAPEGSVIATYMSFSI:POLVLEAEAEVMAEERVVM 180
Db      121 KDAKLKLYSGVPLGPHKESAVTAPEGSVIATYMSFSI:POLVLEAEAEVMAEERVVM 180
Qy      67  -----LVLEAEAEVMAEERVVM 82
Db      67  -----LVLEAEAEVMAEERVVM 82
Qy      181 LPPARSLKSFVYTSVVAAPFTDSTKVORTQDNCSFGI:HAAGVELMRAFTTGPDPSPYPA 240
Db      181 LPPARSLKSFVYTSVVAAPFTDSTKVORTQDNCSFGI:HAAGVELMRAFTTGPDPSPYPA 240
Qy      83 LPPARSLKSFVYTSVVAAPFTDSTKVORTQDNCSFGI:HAAGVELMRAFTTGPDPSPYPA 142
Db      83 LPPARSLKSFVYTSVVAAPFTDSTKVORTQDNCSFGI:HAAGVELMRAFTTGPDPSPYPA 142
Qy      241 HARCQMLRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLS:PMERHALVOLCGTYPSP 300
Db      241 HARCQMLRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLS:PMERHALVOLCGTYPSP 300
Qy      143 HARCQMLRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLS:PMERHALVOLCGTYPSP 202
Db      143 HARCQMLRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLS:PMERHALVOLCGTYPSP 202
Qy      301 YNLTFSSQVLLITLITNTERRHGFEATFQQLPRMS:CGRLRKAQGTFSPPYPGHY 360
Db      301 YNLTFSSQVLLITLITNTERRHGFEATFQQLPRMS:CGRLRKAQGTFSPPYPGHY 360
Qy      203 YNLTFSSQVLLITLITNTERRHGFEATFQQLPRMS:CGRLRKAQGTFSPPYPGHY 262
Db      203 YNLTFSSQVLLITLITNTERRHGFEATFQQLPRMS:CGRLRKAQGTFSPPYPGHY 262
Qy      361 PENIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPKDYVE:INGEKYCGERSQFVYTS 420
Db      361 PENIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPKDYVE:INGEKYCGERSQFVYTS 420
Qy      263 PENIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPKDYVE:INGEKYCGERSQFVYTS 322
Db      263 PENIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPKDYVE:INGEKYCGERSQFVYTS 322
Qy      421 NSNKTIVRFHSDSYTDGTGFLAEVLSYDSSDPCPGQFTCR:GRCIRKELRCDMADCTDH 480
Db      421 NSNKTIVRFHSDSYTDGTGFLAEVLSYDSSDPCPGQFTCR:GRCIRKELRCDMADCTDH 480

```

Db 323 NSNKITVRFHSDSYDTGFLAEYLSYSSDPCGQFTCRGCRIRKELRCDGMADCTDH 382  
Qy 481 SDELNCSCDAGHOFTCKNFKCKPLFWVCDSVNDGSDSGSCPAQOTFRCSNGKCLSK 540  
Db 383 SDELNCSCDAGHOFTCKNFKCKPLFWVCDSVNDGSDSGSCPAQOTFRCSNGKCLSK 442  
Qy 541 SQQCNKDCDGDSDSASCPKRVNVVTCTKHYRCLNGLCLSKNPECDGKEDSDGSD 600  
Db 443 SQQCNKDCDGDSDSASCPKRVNVVTCTKHYRCLNGLCLSKNPECDGKEDSDGSD 502  
Qy 601 DCCGGLASFTRQARVVGCTADBEEMWQVSLHALGQGHICGASLISPMNLVSAACYID 660  
Db 503 DCCGGLASFTRQARVVGCTADBEEMWQVSLHALGQGHICGASLISPMNLVSAACYID 562  
Qy 661 DRGFRYSDPTOMTAFGLHDQSORAPGVQERLKRITISHPFNDFTFYDIALLELEXP 720  
Db 563 DRGFRYSDPTOMTAFGLHDQSORAPGVQERLKRITISHPFNDFTFYDIALLELEXP 622  
Qy 721 AEYSSWVRPICLPDASHVPAGKAIWVTGMGHTQYGGTGAIIQKEIRIVINOTTCENLL 780  
Db 623 AEYSSWVRPICLPDASHVPAGKAIWVTGMGHTQYGGTGAIIQKEIRIVINOTTCENLL 682  
Qy 781 PQOTPRMNCVGLSGVDSGQDSSGGLSSVEADGRIFGAGVYVSWGDCAGRNKPGVYT 840  
Db 683 PQOTPRMNCVGLSGVDSGQDSSGGLSSVEADGRIFGAGVYVSWGDCAGRNKPGVYT 742  
Qy 841 RLPLFRDWIKENTGV 855  
Db 743 RLPLFRDWIKENTGV 757

RESULT 26  
US-09-900-751-2  
; Sequence 2, Application US/09900751  
; Patent No. US20020026653A1  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Keith D  
; APPLICANT: Leviten, Michael W.  
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING SERINE  
; TITLE OF INVENTION: PROTEASE GENE DISRUPTIONS  
; FILE REFERENCE: R-386  
; CURRENT APPLICATION NUMBER: US/09/900,751  
; CURRENT FILING DATE: 2001-07-06  
; PRIOR APPLICATION NUMBER: US 60/217,449  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: US 60/223,170  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: US 60/223,460  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-900-751-2

Query Match 82.9%; Score 3884; DB 9; Length 855;  
Best Local Similarity 81.5%; Pred. No. 2e-288;  
Matches 697; Conservative 73; Mismatches 85; Indels 0; Gaps 0;  
Qy 1 MGSDDRKGGGPGPDGAGLKYNRHEKVNGLBEGVFLPVNNVKKYKKGPRWVTLAA 60  
Db 1 MGSNRGRKAGGQDPAAGLKYNRHEKVNGLBEGVFLPVNNVKKYKKGPRWVTLAA 60  
Qy 61 VILGLLVILGIGFLVHLQYRDYVQVQKXGKIRITNENFVDAYENSNSTEPYSLASKV 120  
Db 61 VLFPSFLSLIMAGLIVHMFYRNRRVQKVFNGHLRITNEIFLDAYENSTSTEFISLASQV 120  
Qy 121 KDLAKLISYGVPLGPGYHKSAAVTAPEGSVIAIYVSEFSIPQILVFEARVNAEEVYM 180  
Db 121 KEALKLISYGVPLGPGYHKSAAVTAPEGSVIAIYVSEFSIPQILVFEARVNAEEVYM 180

Qy 181 LPPRASILKSFVTVSVAPPTDSKTQVORTQDSCSFGHARGVELMRFTTPGPDSPPYA 240  
Db 181 LPPRASILKSFVTVSVAPPTDSKTQVORTQDSCSFGHARGVELMRFTTPGPDSPPYA 240  
Qy 241 HARCQWLRKDDADSVLSLFRSPDLASCDERGSIDLTVYVNTLSPPMPPHALVQICGYPPS 300  
Db 241 HARCQWLRKDDADSVLSLFRSPDLASCDERGSIDLTVYVNTLSPPMPPHALVQICGYPPS 300  
Qy 301 YNLTFFSSQNVLLITLITVTERHPGFATFPOLPMSSCGGLRAQGTFPNSPPYPGHY 360  
Db 301 YNLTFFSSQNVLLITLITVTERHPGFATFPOLPMSSCGGLRAQGTFPNSPPYPGHY 360  
Qy 361 PENIDTWNIEVNNQHVNRKFFYLLBPGVAGTCPCPDYVINEKTYCGERSQFVYTS 420  
Db 361 PENIDTWNIEVNNQHVNRKFFYLLBPGVAGTCPCPDYVINEKTYCGERSQFVYTS 420  
Qy 421 NSNKITVRFHSDSYDTGFLAEYLSYSSDPCGQFTCRGCRIRKELRCDGMADCTDH 480  
Db 421 NSNKITVRFHSDSYDTGFLAEYLSYSSDPCGQFTCRGCRIRKELRCDGMADCTDH 480  
Qy 481 SDELNCSCDAGHOFTCKNFKCKPLFWVCDSVNDGSDSGSCPAQOTFRCSNGKCLSK 540  
Db 481 SDELNCSCDAGHOFTCKNFKCKPLFWVCDSVNDGSDSGSCPAQOTFRCSNGKCLSK 540  
Qy 541 SQQCNKDCDGDSDSASCPKRVNVVTCTKHYRCLNGLCLSKNPECDGKEDSDGSD 600  
Db 541 SQQCNKDCDGDSDSASCPKRVNVVTCTKHYRCLNGLCLSKNPECDGKEDSDGSD 600  
Qy 601 DCCGGLASFTRQARVVGCTADBEEMWQVSLHALGQGHICGASLISPMNLVSAACYID 660  
Db 601 DCCGGLASFTRQARVVGCTADBEEMWQVSLHALGQGHICGASLISPMNLVSAACYID 660  
Qy 661 DRGFRYSDPTOMTAFGLHDQSORAPGVQERLKRITISHPFNDFTFYDIALLELEXP 720  
Db 661 DRGFRYSDPTOMTAFGLHDQSORAPGVQERLKRITISHPFNDFTFYDIALLELEXP 720  
Qy 721 AEYSSWVRPICLPDASHVPAGKAIWVTGMGHTQYGGTGAIIQKEIRIVINOTTCENLL 780  
Db 721 AEYSSWVRPICLPDASHVPAGKAIWVTGMGHTQYGGTGAIIQKEIRIVINOTTCENLL 780  
Qy 781 PQOTPRMNCVGLSGVDSGQDSSGGLSSVEADGRIFGAGVYVSWGDCAGRNKPGVYT 840  
Db 781 PQOTPRMNCVGLSGVDSGQDSSGGLSSVEADGRIFGAGVYVSWGDCAGRNKPGVYT 840  
Qy 841 RLPLFRDWIKENTGV 855  
Db 841 RLPLFRDWIKENTGV 855

RESULT 27  
US-10-072-012-355  
; Sequence 355, Application US/10072012  
; Publication No. US20040003493A1  
; GENERAL INFORMATION:  
; APPLICANT: Tchernev, Velizar  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Patuturajan, Meera  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Li, Li  
; APPLICANT: Gangolli, Baha  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Anderson, David W.  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Taupier Jr, Raymond J.  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Colman, Steven D.  
; APPLICANT: Wolenc, Carol E. A  
; APPLICANT: Bena, Carol E. A  
; APPLICANT: Furtak, Katarzyna

APPLICANT: Groose, William M.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Lepley, Denise M.  
APPLICANT: Rieger, Daniel K.  
APPLICANT: Burgess, Catherine E.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-258  
CURRENT APPLICATION NUMBER: US/10/072,012  
CURRENT FILING DATE: 2002-01-31  
PRIOR APPLICATION NUMBER: 60/265,102  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: 60/265,514  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,517  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,412  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,395  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/266,406  
PRIOR FILING DATE: 2001-02-02  
PRIOR APPLICATION NUMBER: 60/266,767  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: 60/267,057  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/266,975  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/267,459  
PRIOR FILING DATE: 2001-02-08  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1391  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 355  
LENGTH: 855  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-072-012-355

Query Match 82.9%; Score 3884; DB 15; Length 855;  
Best Local Similarity 81.5%; Pred. No. 2e-288;  
Matches 697; Conservative 73; Mismatches 85; Indels 0; Gaps 0;

1 MGSDRRKGGGKPDGAGLKYNSRHEKVLGEVEFLPVNNVKKYKKGPRHVVLLAA 60  
1 MGSRRGKAGGSGDFGAGLKYNSRLNEMGFEEGVEFLPVNNAKYKKGPRHVVLLAA 60  
61 VLIGLLVLGIGFLVHLYQYRDVYVQKXKXNRYITNENFVDAYENSNSTEFLASKV 120  
61 VLFSFLLSLMAGLLVHFFHRYNRVQKVFNGHLRTNIEFLDAYENSTSTEFISLSQV 120  
121 KDALKLISGVPLGPIPHKESAVTAFSEGSYIATYWSFESIPOHLVEAEKVMAEEVVM 180  
121 KEALKLYNEVPVIGPYHKSAVTAFSEGSYIATYWSFESIPOHLVEAEKVMAEEVVM 180  
181 LPPRARSLSKSPVTVSVAFPTDSKTVORTODNSCFGHARGVLMPTTGPSPSPYPA 240  
181 LPPRARKLSKSVTLTSVAFPTDPMLORTODNSCFGHARGVLMPTTGPSPSPYPA 240  
241 HARCQMLRGDADSVLSTFRSPDLASCDERGSDLVTYNTLSPEMHPALVOLCGTYP 300  
241 HARCQMLRGDADSVLSTFRSPDLASCDERGSDLVTYNTLSPEMHPALVOLCGTYP 300  
301 YNLTFHSSQNVLLTLTNTERRRPGFEATFPOLPRNNSCGGRARAKQGTNSPYRGY 360  
301 YNLTFHSSQNVLLTLTNTERRRPGFEATFPOLPRNNSCGGRARAKQGTNSPYRGY 360  
361 PPNIDCTWNTIEPVNNQVYKFKFFYLLEPGVPATGCEKDYVEINGEKCYCGERSQFV 420  
361 PPNINCTWNTIEPVNNQVYKFKFFYLLEPGVPATGCEKDYVEINGEKCYCGERSQFV 420  
421 NSNKITVRFHSDQSYTDTGLAEVLYSDSDPCGQFTCRGTGRICRKLRCGMACTDH 480  
421 NSNKITVRFHSDQSYTDTGLAEVLYSDSDPCGQFTCRGTGRICRKLRCGMACTDH 480

481 SDELNCSGADAGHPTCKNFKCKPLFWYCDVNDGSDNDEGSCCPAQTRFCSNGKCLSK 540  
481 SDERYCKNAHTHQTCKNQCKPLFWYCDVNDGSDNDEGSCCPAGSKCSNGKCLPQ 540  
541 SQQNGKDDGSDSDSASCPKVVVYVTKTHTYCLNGLCLSKNPECDEKSDGSD 600  
541 SQKNGKDDGSDSDSASCPKVVVYVTKTHTYCLNGLCLSKNPECDEKSDGSD 600  
601 DCCGGLSFTROARVVGTDADGEMPMQVSLHALGSGHICGASLISPMNLVSAHCT 660  
601 NCCGGLSFTROARVVGTDADGEMPMQVSLHALGSGHICGASLISPMNLVSAHCT 660  
661 DRGFRSDPTQWPAFLGLHDOSASAGYVERLKRISHPFNDFTFDDIALLEKXP 720  
661 DKRFKSDYTMWPAFLGLHDOSASAGYVERLKRISHPFNDFTFDDIALLEKXS 720  
721 AEVSSWVRPCLDPASHVPAGRAIWTGHTQYGGTGLLIQKGEIRVINOCTCENLL 780  
721 VEYSTVVRPCLDPATHVPAGRAIWTGHTKGGTGLLIQKGEIRVINOCTCEDLM 780  
781 PQQITPRMCCVGRISGCVDSGCGDGGPLSSVEADGHIIFAGVYWSGDCAGRNKPGV 840  
781 PQQITPRMCCVGRISGCVDSGCGDGGPLSSVEADGHIIFAGVYWSGDCAGRNKPGV 840  
841 RLPLFRMKEKNTGV 855  
841 RLPLFRMKEKNTGV 855

RESULT 28  
US-10-072-012-413  
Sequence 413, Application US/10072012  
Publication No. US20040033493A1  
GENERAL INFORMATION:  
APPLICANT: Tcherev, Velizar  
APPLICANT: Spytek, Kimberly  
APPLICANT: Zethusen, Bryan  
APPLICANT: Patturajan, Meera  
APPLICANT: Shinkets, Richard  
APPLICANT: Li, Li  
APPLICANT: Gangoli, Esha  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Anderson, David W.  
APPLICANT: Rastelli, Luca  
APPLICANT: Miller, Charles E.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Taupier Jr, Raymond J.  
APPLICANT: Gusev, Vladimir Y.  
APPLICANT: Coleman, Steven D.  
APPLICANT: Wolenc, Adam R.  
APPLICANT: Pena, Carol E. A  
APPLICANT: Furtak, Katarzyna  
APPLICANT: Groose, William M.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Lepley, Denise M.  
APPLICANT: Rieger, Daniel K.  
APPLICANT: Burgess, Catherine E.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-258  
CURRENT APPLICATION NUMBER: US/10/072,012  
CURRENT FILING DATE: 2002-01-31  
PRIOR APPLICATION NUMBER: 60/265,102  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: 60/265,514  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,517  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,412  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,395  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/266,406

```

; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 413
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-072-012-413

Query Match      82.9%; Score 3884; DB 15; Length 855;
Best Local Similarity 81.5%; Pred. No. 2e-288;
Matches 697; Conservative 73; Mismatches 85; Indels 0; Gaps 0;

Qy      1  MGSDBARKGGGKPDGAGLKYSRHKXNGLEGVFLVNNVKKYKPKGKRWVLA 60
Db      1  MGSNRGRKAGGSDPGAGLKYSRLNNMNGFEEGVFLPANNAKYKPKRWVLA 60

Qy      61  VLIGLLVLIGLIGFLVHLYRDRVQKXGKGYMRITNENFVDAYENSNGTEFVSLASKV 120
Db      61  VLFSFLLSLMAGLLVHMFYRNVQKXVNGHRIKINELFLDAYENSTSTEFVSLASKV 120

Qy      121  KDALKLIYSGVPLGPRYHKESAVTAFSEGSVIAYWSEFSIPOHLYEAEERVMAEERV 180
Db      121  KEALKLILYNEVPVLYGPYHKKSAVTAFSEGSVIAYWSEFSIPIHLAEVDRAMAEV 180

Qy      181  LPPRARSLSKFVTVSVVAFPTDSKTQVORTODNSCFELHARGVLEMTTGPFPDSPY 240
Db      181  LPPRARRLSKFLSVVAFPTDPMLOKTDNSCFELHARGVAVTRFTTGPFPNSPY 240

Qy      241  HARCOWALRGDADSVLSLTFRSPDLASCDERGSGLVTVYNTLSMEPHALVOLCGT 300
Db      241  HARCOWALRGDADSVLSLTFRSPDLASCDERGSGLVTVYNTLSMEPHALVOLCGT 300

Qy      301  YNLTFFSSQVNLTLITNTERHHPGFEATFPLPMSSCGGRIRKAQGTFSNBPYRG 360
Db      301  YNLTFFSSQVNLTLITNTERHHPGFEATFPLPMSSCGGRIRKAQGTFSNBPYRG 360

Qy      361  PPNIDCWNIENVEPNNOVKVRFKFEVLYLEBGPVAGTCPKQVYELNGEKYGEBSQ 420
Db      361  PPNIDCWNIENVEPNNOVKVRFKFEVLYLEBGPVAGTCPKQVYELNGEKYGEBSQ 420

Qy      421  NSNKITVRFHSDSYTDTGFLAEYLSYSSDPCGQFTCTGRCIRKELRCDDGADCT 480
Db      421  NSNKITVRFHSDSYTDTGFLAEYLSYSSDPCGQFTCTGRCIRKELRCDDGADCT 480

Qy      481  SDELINSCDAGHQTCKNCKCKPLFWYCDVNDGDSDBQSGSCPAQTRCSNGKLSK 540
Db      481  SDRRYRCNATHTQTCNQCFCKPLFWYCDVNDGDSDBQSGSCPAQTRCSNGKLSK 540

Qy      541  SQQONGKDDGSDSDEASCPKAVVVTCTKHTYRCLNGLCLSKNPECDGKEDSDG 600
Db      541  SQQONGKDDGSDSDEASCPKAVVVTCTKHTYRCLNGLCLSKNPECDGKEDSDG 600

Qy      601  DDCGGLSFTROARVVGTDADDEGEPMQVSLHALGSHICGASLSPNMLVSAHAY 660
Db      601  NCCGGLSFTROARVVGTDADDEGEPMQVSLHALGSHICGASLSPNMLVSAHAY 660

Qy      661  DRGFRISDPTQWTAFLGLHDQSORAPGVVERRLKRIISHPFNDFFPDYDIALLE 720
Db      661  DRGFRISDPTQWTAFLGLHDQSORAPGVVERRLKRIISHPFNDFFPDYDIALLE 720

Qy      721  AEXYSMWPRICLDASHVFPAGKAIWYTGNGHNYGSGTGLILIOKGIIRVINTOT 780
Db      721  VETSTVVRPICLPDATHVFPAGKAIWYTGNGHNYGSGTGLILIOKGIIRVINTOT 780
```

```

Qy      781  PQQITPRMVCVGLSGGVDSQQDSGAPLSVYADGRIFGAGVYVSWGDCAGRNK 840
Db      781  PQQITPRMVCVGLSGGVDSQQDSGAPLSVYADGRIFGAGVYVSWGDCAGRNK 840

Qy      841  RLPVFRDWIKENTGV 855
Db      841  RLPVFRDWIKENTGV 855

RESULT 29
US-10-072-012-356
; Sequence 356, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Baha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Molenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 356
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-072-012-356

Query Match      82.6%; Score 3866; DB 15; Length 855;
Best Local Similarity 80.8%; Pred. No. 4.7e-287;
```

[illegible]

RESULT 30  
US-10-072-012-414  
Sequence 414, Application US/10072012  
Publication No. US20040033493A1  
GENERAL INFORMATION:  
APPLICANT: Tchernev, Velizar  
APPLICANT: Spivey, Kimberly  
APPLICANT: Zernusen, Bryan  
APPLICANT: Patursajan, Meera  
APPLICANT: Shukhats, Richard

```

APPLICANT: Li, Li
APPLICANT: Gangolli, Esna
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Colman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grosse, William M.
APPLICANT: Alabrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 414
LENGTH: 855
TYPE: PRT
ORGANISM: Rattus norvegicus
US-10-072-012-414

```

[illegible]

```

Db      241 HARCQWVLRGDADSVLSLTRSFVAPCDGHDSDLVTVYDLSLSPMEPHAVVRLCGTSPS 300
QY      301 YNLTFHSSQNVLLITLLINTERRHPGEATFFQLPRMSSCGRLRKAQGTENSPYPGNY 360
Db      301 YNLTFHSSQNVFLVTLITNDRHRHGFEPALFFQLPKMSSCGGLSEAGTFSPPYPGHY 360
QY      361 PPNIDCTWNIIEVNNQVKKRFEKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVYTS 420
Db      361 PPNINCTWNIKVNNRNVKVRFKLFYLVDPNI PVGSC TKDYVEINGEKFCGERSQFVYSS 420
QY      421 NSNKITVRFHSDOSYTDGTGFLAEYLSYSDSPCGOFTCRTRGCRIRKELRCDGWADCTDH 480
Db      421 NSSKITVHFHSDHSYTTDTGFLAEYLSYSDNDPCGFMWCKTGRCIRKDLRCDGMADCPDY 480
QY      481 SDELRCSGADAGHOFTCKNKECKPLFWVCDSVNDGSDNSDEQCSCPAOTFRCSNGKCLSK 540
Db      481 SDERHCRGNATHQFMCKNQFCCKPLFWVCDSVNDGSDSDEGCSCPAGSFKCSNGKCLPQ 540
QY      541 SOQCNKGDDCGSDSDASCPKRVNVYTCYKTYRCLNGLCLSKGNPECDGKEDCGSDSEK 600
Db      541 SOQCNKGDDCGSDSDASCDNVNAVSCYKTYRCONGLCLNKGNPECDGKEDCGSDSEK 600
QY      601 DCDGGLRSFTROARVVGTDADGEMPMOYSLHALGQGHICGASLISPNMLVSAHGYID 660
Db      601 NCDGGLRSFTKQARVVGSTNADGEMPMOYSLHALGQGHICGASLISPDMLVSAHCFOD 660
QY      661 DRGFRYSDPQTQWTAFLGLHDQSORABGVQERLKRILSHPFNDFTFYDIALLELEKP 720
Db      661 ETIFKYSDDHTMTAFGLLDQSKRSASGVQEHKLRILITHPSFNDFTFYDIALLELEKP 720
QY      721 AEYSSMWRPICLPDASHVPFPAKALIMYTGNGHTOYGGTGALLQKGEIRVINOTTGENTL 780
Db      721 AEYSTVVRPICLPDNTHVFPAGKALIMYTGNGHTKEGGTGALLQKGEIRVINOTTCELL 780
QY      781 PQOITPRMNCVGFISGVSDSCQSDGGLPSSVEADGRIFGAGVYWSGDGCAGRNKPQVYT 840
Db      781 PQOITPRMNCVGFISGVSDSCQSDGGLPSSVEKDGRIFGAGVYWSGEGCAQRNKPQVYT 840
QY      841 RLPPLFRDIIKENTGV 855
Db      841 RIPEVRDIIKEQOTGV 855

```

Search completed: November 29, 2004, 08:49:57  
 Job time : 118.183 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: November 29, 2004, 08:20:59 : Search time 32.2432 Seconds  
(without alignments)  
1758.570 Million cell updates/sec

Title: US-09-936-333-27  
Perfect score: 4683  
Sequence: 1 MGSDRARKGGCGPDPAGL.....PGVYTRLPFRDWIKENTGV 855

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 6618000 residues  
Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :  
1: /cgcn2\_6/ptodata/1/1aa/5A.COMB.pdp:\*  
2: /cgcn2\_6/ptodata/1/1aa/5B.COMB.pdp:\*  
3: /cgcn2\_6/ptodata/1/1aa/6A.COMB.pdp:\*  
4: /cgcn2\_6/ptodata/1/1aa/6B.COMB.pdp:\*  
5: /cgcn2\_6/ptodata/1/1aa/PCITUS.COMB.pdp:\*  
6: /cgcn2\_6/ptodata/1/1aa/backfile1.pdp:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4653	99.4	855	2	US-09-027-337-2
2	4653	99.4	855	4	US-09-644-600-2
3	4653	99.4	855	4	US-09-654-600A-2
4	3793	81.0	902	4	US-09-644-600-10
5	3793	81.0	902	4	US-09-654-600A-10
6	1305	27.9	241	4	US-09-657-986B-2
7	700.5	15.0	798	1	US-08-200-900A-2
8	700.5	15.0	798	4	US-08-794-042-2
9	700.5	15.0	798	5	PCT-US94-00616-2
10	583	12.4	407	3	US-09-734-675-4
11	557.5	11.9	492	4	US-09-685-166A-895
12	557.5	11.9	492	4	US-09-879-792-14
13	557.5	11.9	492	4	US-09-679-426-895
14	555.5	11.9	492	4	US-09-342-749-2
15	555.5	11.9	492	4	US-09-641-840-2
16	543.5	11.6	235	3	US-08-944-483-65
17	540.5	11.5	235	3	US-08-807-151-3
18	540.5	11.5	235	3	US-09-478-957-3
19	536.5	11.5	454	3	US-09-518-046-2
20	518	11.1	446	4	US-10-177-661-4
21	517.5	11.1	418	4	US-10-177-661-6
22	509.5	10.9	283	3	US-08-807-151-1
23	509.5	10.9	283	3	US-09-478-957-1
24	509.5	10.9	477	4	US-10-177-661-2
25	509.5	10.9	562	4	US-09-879-792-12
26	508.5	10.9	418	1	US-08-508-448C-25
27	508.5	10.9	418	4	US-09-370-838-82

28	508.5	10.9	418	4	US-09-370-838-83	Sequence 83, Appl
29	508.5	10.9	418	4	US-09-854-133-82	Sequence 82, Appl
30	508.5	10.9	418	4	US-09-854-133-83	Sequence 83, Appl
31	507.5	10.8	418	4	US-09-370-838-82	Sequence 82, Appl
32	507.5	10.8	418	4	US-09-854-133-82	Sequence 82, Appl
33	507.5	10.8	638	2	US-08-681-151-3	Sequence 3, Appl
34	506.5	10.8	790	4	US-08-991-761A-13	Sequence 13, Appl
35	501.5	10.7	376	4	US-09-830-002-2	Sequence 2, Appl
36	494.5	10.6	232	2	US-08-508-448C-19	Sequence 19, Appl
37	492	10.5	516	2	US-09-000-846-2	Sequence 2, Appl
38	491.5	10.5	546	6	5200340-6	Patent No. 5200340
39	489	10.4	791	1	US-08-643-219-1	Sequence 1, Appl
40	489	10.4	791	3	US-08-851-350-1	Sequence 1, Appl
41	488.5	10.4	583	4	US-09-976-594-837	Sequence 837, App
42	488	10.4	417	4	US-09-820-002-4	Sequence 4, Appl
43	487	10.4	791	2	US-09-131-995-1	Sequence 1, Appl
44	487	10.4	791	2	US-08-832-087B-1	Sequence 1, Appl
45	487	10.4	791	3	US-09-132-154-1	Sequence 1, Appl
46	487	10.4	791	4	US-08-991-761A-6	Sequence 6, Appl
47	487	10.4	791	4	US-08-924-287A-1	Sequence 1, Appl
48	487	10.4	810	1	US-07-854-603-2	Sequence 2, Appl
49	487	10.4	810	1	US-08-147-000B-29	Sequence 29, Appl
50	487	10.4	810	3	US-09-086-514-1	Sequence 1, Appl
51	487	10.4	810	4	US-09-132-012-5	Sequence 5, Appl
52	487	10.4	810	4	US-09-403-736-1	Sequence 1, Appl
53	487	10.4	810	4	US-09-701-265-1	Sequence 1, Appl
54	487	10.4	810	6	5200340-8	Patent No. 5200340
55	485.5	10.4	812	4	US-08-991-761A-7	Sequence 7, Appl
56	484.5	10.3	790	1	US-08-469-486-54	Sequence 54, Appl
57	484.5	10.3	790	2	US-08-469-558-54	Sequence 54, Appl
58	484	10.3	812	1	US-08-750-711-1	Sequence 1, Appl
59	483.5	10.3	812	1	US-08-248-659A-1	Sequence 1, Appl
60	483.5	10.3	812	1	US-08-452-260-1	Sequence 1, Appl
61	483.5	10.3	812	1	US-08-452-260-1	Sequence 1, Appl
62	483.5	10.3	812	1	US-08-326-785-1	Sequence 1, Appl
63	483.5	10.3	812	2	US-08-612-788-1	Sequence 1, Appl
64	483.5	10.3	812	2	US-08-605-598B-1	Sequence 1, Appl
65	483.5	10.3	812	2	US-08-429-743-1	Sequence 1, Appl
66	483.5	10.3	812	2	US-08-866-735-1	Sequence 1, Appl
67	483.5	10.3	812	3	US-09-066-028-1	Sequence 1, Appl
68	483.5	10.3	812	4	US-09-192-012-3	Sequence 3, Appl
69	483.5	10.3	812	4	US-09-335-325-1	Sequence 1, Appl
70	483.5	10.3	812	4	US-08-991-761A-12	Sequence 12, Appl
71	483.5	10.3	812	5	PCT-US95-05107-1	Sequence 1, Appl
72	479.5	10.2	415	1	US-08-073-531B-1	Sequence 1, Appl
73	479.5	10.2	415	1	US-08-766-288-1	Sequence 1, Appl
74	477	10.2	810	4	US-08-991-761A-11	Sequence 11, Appl
75	475	10.1	328	4	US-09-386-642-11	Sequence 11, Appl
76	474.5	10.1	256	2	US-09-027-337-3	Sequence 3, Appl
77	474.5	10.1	256	4	US-09-644-600-3	Sequence 3, Appl
78	474.5	10.1	256	4	US-09-654-600A-3	Sequence 3, Appl
79	474.5	10.1	415	1	US-08-295-411-2	Sequence 2, Appl
80	474.5	10.1	415	5	US-08-955-471-2	Sequence 2, Appl
81	474.5	10.1	415	5	PCT-US92-10242-2	Sequence 2, Appl
82	474.5	10.1	461	3	US-08-742-877-2	Sequence 2, Appl
83	472	10.1	405	3	US-09-734-675-2	Sequence 2, Appl
84	472	10.1	461	6	5521070-2	Patent No. 5521070
85	471.5	10.1	315	4	US-09-386-653A-9	Sequence 9, Appl
86	471	10.1	406	4	US-09-851-588-6	Sequence 6, Appl
87	471	10.1	423	4	US-09-656-002-2	Sequence 2, Appl
88	471	10.1	435	3	US-09-008-271A-6	Sequence 6, Appl
89	471	10.1	435	4	US-09-607-745-2	Sequence 2, Appl
90	471	10.1	437	4	US-09-851-588-8	Sequence 8, Appl
91	471	10.1	461	3	US-09-033-871A-21	Sequence 21, Appl
92	471	10.1	461	4	US-10-133-907-5	Sequence 5, Appl
93	469.5	10.0	255	3	US-08-944-483-67	Sequence 6, Appl
94	469	10.0	319	4	US-09-386-642-12	Sequence 12, Appl
95	468.5	10.0	415	4	US-09-118-748-2	Sequence 2, Appl
96	468	10.0	455	3	US-09-261-416-2	Sequence 2, Appl
97	468	10.0	527	6	5520913-1	Patent No. 5520913
98	466.5	10.0	290	4	US-09-386-653A-7	Sequence 7, Appl
99	466.5	10.0	290	4	US-10-140-002-222	Sequence 222, App
100	463	9.9	527	2	US-08-811-949-39	Sequence 39, Appl

ALIGNMENTS

```
RESULT 1
US-09-027-337-2
; Sequence 2, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: Tanimoto, Timothy J.
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides
; Patent No. 5972616
US-09-027-337-2

Query Match      99.4%; Score 4653; DB 2; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKVNGLSEGVFLPVNNVKYKKGPGRMVLLAA 60
DB 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKVNGLSEGVFLPVNNVKYKKGPGRMVLLAA 60
QY 61 VLIGLLVLLIGLIFLVNHLQYRDVAVQKXNGYRNTNENFVDAYENSNSTEYVLSASKV 120
DB 61 VLIGLLVLLIGLIFLVNHLQYRDVAVQKXNGYRNTNENFVDAYENSNSTEYVLSASKV 120
QY 121 KDLAKLLYSQVPLGPGYHKSAAVTAFSEGSVIAYYSEFSIPQHLVEAEKVAEERVVM 180
DB 121 KDLAKLLYSQVPLGPGYHKSAAVTAFSEGSVIAYYSEFSIPQHLVEAEKVAEERVVM 180
QY 181 LPPRARSLSKFVVTSVVAFTPDSKTVOQTODNSCSFGIHAAGVLMRFTTGPSPSPYPA 240
DB 181 LPPRARSLSKFVVTSVVAFTPDSKTVOQTODNSCSFGIHAAGVLMRFTTGPSPSPYPA 240
QY 241 HARCQALRGDADSVSLTFRSPDLASCDERGSDLVTYNTTSLSPMEPHALVOLCGTYPSP 300
DB 241 HARCQALRGDADSVSLTFRSPDLASCDERGSDLVTYNTTSLSPMEPHALVOLCGTYPSP 300
QY 301 YNLTFFSSQNVLLITLTNTERRHHPGEATFPQLPRMSSCGRLRKAQGTFNSPYTGHY 360
DB 301 YNLTFFSSQNVLLITLTNTERRHHPGEATFPQLPRMSSCGRLRKAQGTFNSPYTGHY 360
QY 361 PPNIDCTWNIEVPNNQHVAFKFEFYLLBEPVPAATCPKDYVEINGEKYCGERSQFVVT 420
DB 361 PPNIDCTWNIEVPNNQHVAFKFEFYLLBEPVPAATCPKDYVEINGEKYCGERSQFVVT 420
QY 421 NSNKITVRFHSDDSYTDTGLAEYLSYDSDPCPGQGTCTGTGRCIRKELRCDGMACTDH 480
DB 421 NSNKITVRFHSDDSYTDTGLAEYLSYDSDPCPGQGTCTGTGRCIRKELRCDGMACTDH 480
QY 481 SDELNGSCDAGHOTCTCKNFKCPFLFWVCDSVNDGSDSCGSPAQOTFRCSNGKCLSK 540
DB 481 SDELNGSCDAGHOTCTCKNFKCPFLFWVCDSVNDGSDSCGSPAQOTFRCSNGKCLSK 540
QY 541 SQQCNGKDDGSDGDEASCPKVVNVCTKHTYRCLNGLCLSKNPECDGKEDCSDGSDK 600
DB 541 SQQCNGKDDGSDGDEASCPKVVNVCTKHTYRCLNGLCLSKNPECDGKEDCSDGSDK 600
QY 601 DCDGCLRSFTFRQARVVGCTADBEQWVQVSLHALGGGHTCGASLISPNMLVSAAHCYID 660
```

```
DB 601 DCDGCLRSFTFRQARVVGCTADBEQWVQVSLHALGGGHTCGASLISPNMLVSAAHCYID 660
QY 661 DRGFRYSDDPTQMTAPLGIHQDSORSA PGVERLRKLIISHPFNDPFEDVILLLEKXP 720
DB 661 DRGFRYSDDPTQMTAPLGIHQDSORSA PGVERLRKLIISHPFNDPFEDVILLLEKXP 720
QY 721 AEYSMWVPICLPDASHVPPAGKAIWYTGHTQYGGTGLILQKEIRVYINQTCENLL 780
DB 721 AEYSMWVPICLPDASHVPPAGKAIWYTGHTQYGGTGLILQKEIRVYINQTCENLL 780
QY 781 PQOITPRMVCVGLSGVDSCQDSSGPISSVADGRIPAGVYVSGDCAGRNKPGVYT 840
DB 781 PQOITPRMVCVGLSGVDSCQDSSGPISSVADGRIPAGVYVSGDCAGRNKPGVYT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855
```

RESULT 2

```
US-09-644-600-2
; Sequence 2, Application US/09644600
; Patent No. 6431500
; GENERAL INFORMATION:
; APPLICANT: Tanimoto, Timothy J.
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-2

Query Match      99.4%; Score 4653; DB 4; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKVNGLSEGVFLPVNNVKYKKGPGRMVLLAA 60
DB 1 MGSDBARKGGGGPDPFGAGLKYNSRHEKVNGLSEGVFLPVNNVKYKKGPGRMVLLAA 60
QY 61 VLIGLLVLLIGLIFLVNHLQYRDVAVQKXNGYRNTNENFVDAYENSNSTEYVLSASKV 120
DB 61 VLIGLLVLLIGLIFLVNHLQYRDVAVQKXNGYRNTNENFVDAYENSNSTEYVLSASKV 120
QY 121 KDLAKLLYSQVPLGPGYHKSAAVTAFSEGSVIAYYSEFSIPQHLVEAEKVAEERVVM 180
DB 121 KDLAKLLYSQVPLGPGYHKSAAVTAFSEGSVIAYYSEFSIPQHLVEAEKVAEERVVM 180
QY 181 LPPRARSLSKFVVTSVVAFTPDSKTVOQTODNSCSFGIHAAGVLMRFTTGPSPSPYPA 240
DB 181 LPPRARSLSKFVVTSVVAFTPDSKTVOQTODNSCSFGIHAAGVLMRFTTGPSPSPYPA 240
QY 241 HARCQALRGDADSVSLTFRSPDLASCDERGSDLVTYNTTSLSPMEPHALVOLCGTYPSP 300
DB 241 HARCQALRGDADSVSLTFRSPDLASCDERGSDLVTYNTTSLSPMEPHALVOLCGTYPSP 300
QY 301 YNLTFFSSQNVLLITLTNTERRHHPGEATFPQLPRMSSCGRLRKAQGTFNSPYTGHY 360
DB 301 YNLTFFSSQNVLLITLTNTERRHHPGEATFPQLPRMSSCGRLRKAQGTFNSPYTGHY 360
```

QY 361 PNIDCTWNIIEVNNQHVKVRFKFYYLLBPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420  
DB 361 PNIDCTWNIIEVNNQHVKVSFKFYYLLBPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420  
QY 421 NSNKITVRFHSDOSYTDGTGFLAEYLSYDSDPCPGQFTCRTRGRCIRKELRCDDGADCTDH 480  
DB 421 NSNKITVRFHSDOSYTDGTGFLAEYLSYDSDPCPGQFTCRTRGRCIRKELRCDDGADCTDH 480  
QY 481 SDELINCSADAGHOFCTCKNFKCKPLFWYCDVNDGDNDSDEGSCCPAQTFRCSNGKCLSK 540  
DB 481 SDELINCSADAGHOFCTCKNFKCKPLFWYCDVNDGDNDSDEGSCCPAQTFRCSNGKCLSK 540  
QY 541 SQQNGKDDGSDGSDASCPKVVNVYCTCKHYRCLNGLCLSKNPECDGKEDCSGSDSK 600  
DB 541 SQQNGKDDGSDGSDASCPKVVNVYCTCKHYRCLNGLCLSKNPECDGKEDCSGSDSK 600  
QY 601 DCDGGLRSFTROARVVGCTADDEGEWPMOYSLHALGCGHICGASLISPMNLVSAHACYID 660  
DB 601 DCDGGLRSFTROARVVGCTADDEGEWPMOYSLHALGCGHICGASLISPMNLVSAHACYID 660  
QY 661 DRGFRYSDDPTOMTAFGLLHDOSQSRAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720  
DB 661 DRGFRYSDDPTOMTAFGLLHDOSQSRAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720  
QY 721 AEYSMWRPICLPASHVFPAGKAIWYTGWGHYGGTGALILQKGEIRVYNQTTCCNL 780  
DB 721 AEYSMWRPICLPASHVFPAGKAIWYTGWGHYGGTGALILQKGEIRVYNQTTCCNL 780  
QY 781 PQQITPRMVCVGLSGGVDSQCGDGGPLSSVEADGRIFGAGVSWGDCAGRKPKGYT 840  
DB 781 PQQITPRMVCVGLSGGVDSQCGDGGPLSSVEADGRIFGAGVSWGDCAGRKPKGYT 840  
QY 841 RLPLFRDIXENTGV 855  
DB 841 RLPLFRDIXENTGV 855

RESULT 3  
US-09-654-600A-2  
; Sequence 2, Application US/09654600A  
; Patent No. 6649741  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Tanimoto, Hirotochi  
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
; TITLE OF INVENTION: Overexpressed in Carcinomas  
; FILE REFERENCE: D6064CIP/D  
; CURRENT APPLICATION NUMBER: US/09/654,600A  
; CURRENT FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 09/421,213  
; 09/027,337  
; PRIOR FILING DATE: 1999-10-20  
; 1998-02-20  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: TADG-15  
US-09-654-600A-2

Query Match 99.4%; Score 4653; DB 4; Length 855;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MGSDRARRGGGGPPDFGAGLKYNSRHEKVNGLBEGVEFLPVNNYKVKCKHGGWVYLA 60  
DB 1 MGSDRARRGGGGPPDFGAGLKYNSRHEKVNGLBEGVEFLPVNNYKVKCKHGGWVYLA 60  
QY 61 VLIGLLVLIGLIGFLVHMLQYRDVAVQKVKNGYKMTNENFVDAYENSNSTEFLVSLASKV 120  
DB 61 VLIGLLVLIGLIGFLVHMLQYRDVAVQKVKNGYKMTNENFVDAYENSNSTEFLVSLASKV 120

QY 121 KDALKLISGVPLFGPYHKESAVTAPSEGSVIAYYMSEFSI POHLVEAEERVAEERVM 180  
DB 121 KDALKLISGVPLFGPYHKESAVTAPSEGSVIAYYMSEFSI POHLVEAEERVAEERVM 180  
QY 181 LPPRARSLSFVYTSVVAFPDTSKTYVQRTODNSCSFGLHARGVELMFTTPEGPDSPYA 240  
DB 181 LPPRARSLSFVYTSVVAFPDTSKTYVQRTODNSCSFGLHARGVELMFTTPEGPDSPYA 240  
QY 241 HARCQMALRGDADSVLSLTRSFDLASCDERGSDDLTVYVNTLSPMERHALVOLCGTYPPS 300  
DB 241 HARCQMALRGDADSVLSLTRSFDLASCDERGSDDLTVYVNTLSPMERHALVOLCGTYPPS 300  
QY 301 YNLTFHSSQNVLLITLITNTERHPGFEATFFOLPRMSCGGRLRKAQGTFSNPPYBGHY 360  
DB 301 YNLTFHSSQNVLLITLITNTERHPGFEATFFOLPRMSCGGRLRKAQGTFSNPPYBGHY 360  
QY 361 PNIDCTWNIIEVNNQHVKVRFKFYYLLBPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420  
DB 361 PNIDCTWNIIEVNNQHVKVSFKFYYLLBPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420  
QY 421 NSNKITVRFHSDOSYTDGTGFLAEYLSYDSDPCPGQFTCRTRGRCIRKELRCDDGADCTDH 480  
DB 421 NSNKITVRFHSDOSYTDGTGFLAEYLSYDSDPCPGQFTCRTRGRCIRKELRCDDGADCTDH 480  
QY 481 SDELINCSADAGHOFCTCKNFKCKPLFWYCDVNDGDNDSDEGSCCPAQTFRCSNGKCLSK 540  
DB 481 SDELINCSADAGHOFCTCKNFKCKPLFWYCDVNDGDNDSDEGSCCPAQTFRCSNGKCLSK 540  
QY 541 SQQNGKDDGSDGSDASCPKVVNVYCTCKHYRCLNGLCLSKNPECDGKEDCSGSDSK 600  
DB 541 SQQNGKDDGSDGSDASCPKVVNVYCTCKHYRCLNGLCLSKNPECDGKEDCSGSDSK 600  
QY 601 DCDGGLRSFTROARVVGCTADDEGEWPMOYSLHALGCGHICGASLISPMNLVSAHACYID 660  
DB 601 DCDGGLRSFTROARVVGCTADDEGEWPMOYSLHALGCGHICGASLISPMNLVSAHACYID 660  
QY 661 DRGFRYSDDPTOMTAFGLLHDOSQSRAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720  
DB 661 DRGFRYSDDPTOMTAFGLLHDOSQSRAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720  
QY 721 AEYSMWRPICLPASHVFPAGKAIWYTGWGHYGGTGALILQKGEIRVYNQTTCCNL 780  
DB 721 AEYSMWRPICLPASHVFPAGKAIWYTGWGHYGGTGALILQKGEIRVYNQTTCCNL 780  
QY 781 PQQITPRMVCVGLSGGVDSQCGDGGPLSSVEADGRIFGAGVSWGDCAGRKPKGYT 840  
DB 781 PQQITPRMVCVGLSGGVDSQCGDGGPLSSVEADGRIFGAGVSWGDCAGRKPKGYT 840  
QY 841 RLPLFRDIXENTGV 855  
DB 841 RLPLFRDIXENTGV 855

RESULT 4  
US-09-644-600-10  
; Sequence 10, Application US/09644600  
; Patent No. 6451500  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Tanimoto, Hirotochi  
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
; TITLE OF INVENTION: Overexpressed in Carcinomas  
; FILE REFERENCE: D6064CIP/D  
; CURRENT APPLICATION NUMBER: US/09/644,600  
; CURRENT FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: 09/421,213  
; PRIOR FILING DATE: 1999-10-20  
; PRIOR APPLICATION NUMBER: 09/027,337  
; PRIOR FILING DATE: 1998-02-20  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 10  
; LENGTH: 902

TYPE: PRT  
ORGANISM: Mus musculus  
FEATURE:  
OTHER INFORMATION: Epithin  
US-09-644-600-10

Query Match 81.0%; Score 3793; DB 4; Length 902;  
Best Local Similarity 81.1%; Pred. No. 4,86-269;  
Matches 684; Conservative 71; Mismatches 88; Indels 0; Gaps 0;

```
QY 1 MGSBRARKGKGGPKDGAAGLKYSRHRKVNGLSEGVFLPVNNVKYKXKGPGRWVLLAA 60
D 1 MGSBRARKGKGGKQDPAAGLKYSRLNNMNGFESEVFLPANNAKKXKXKGPGRWVLLVA 60
QY 61 VLIGLLVLIGLIGFLVWHLQYRDVRYOKVKNQYMRITNENFVDAYENSNSTEFVSLASKV 120
D 61 VLSEFLLSLMAGLLVWHFHYRNVRYOKVFNQHLRITNEIFLDAYENSTSTEFVSLASKV 120
QY 121 KDALKLLYSQVPLGPRYHKESAVTAESGVIAIYWSSEFSIPHLYBEAERWMAEERVVM 180
D 121 KEALKLLYNEVPVLGPRYHKESAVTAESGVIAIYWSSEFSIPHLYBEAERWMAEERVVM 180
QY 181 LPPRARSLSFVYTSVVAFPPTDSKTVOPTODNSCSFGLHARGVELMRFTTGPSPSPYPA 240
D 181 LPPRARSLSKSFVTSVVAFPIDPMLQRTODNSCSFALHARGAAYVTRFTTGPSPSPYPA 240
QY 241 HARCQWALRGDADSVLSLTFERSFDLASCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300
D 241 HARCQWALRGDADSVLSLTFERSFDVAPCDEHGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300
QY 301 YNLTFFSSQVNVLLITLITNTERRHHPGEATFFQLPRMSSCGGLRLKAQGTFFNSPYPGHY 360
D 301 YNLTFFLSQVNVFLVTLITNTERRHHPGEATFFQLPRMSSCGGLRLKAQGTFFNSPYPGHY 360
QY 361 PPNIDCTWNTNIEVPNNQHVKKRFEFYLLBPGVAGTCKPDYVEINERKXCGESRQFVYTS 420
D 361 PPNINCTWNTNIEVPNNRNVKRFKLFYLVDPNVVPGSLCKDYVEINERKXGESRQFVYTS 420
QY 421 NSNKTIVRFHSDSYTDTGFLAEYLSYSSDPGPGFTCTGRCIRKELRCDGMADCTDH 480
D 421 NSSKITVYHFHSDSYTDTGFLAEYLSYSSDPGPGFTCTGRCIRKELRCDGMADCTDH 480
QY 481 SDELINCSGADGHQFTCKNFKCKPLFWVCDVNDGSDNSDEQSCCPAQOTFRCSNGKCLSK 540
D 481 SDEYRCRCAVTHQFTCKNFKCKPLFWVCDVNDGSDNSDEQSCCPAGSFKCSNGKCLPQ 540
QY 541 SQQCKNGKDDGSDSDASCPKVVNVVCTKHTYRCLNGLCLSKNPGCDGKEDSDGSDSK 600
D 541 SQQCKNGKDDGSDSDASCDVNVVNSCTKHTYRCLNGLCLSKNPGCDGKEDSDGSDSK 600
QY 601 DCDGGLRSFTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPNMLVSAAHCYID 660
D 601 NCCGGLRSFTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPNMLVSAAHCFOD 660
QY 661 DRGRYSDPTQWTAFLGLHDSQSRAPGVQERRLKRIISHPFNDFTFYDIALLELEKP 720
D 661 DRKFKYSDYTMWTAFLGLHDSQSRASGVQELKRIIHPFNDFTFYDIALLELEKS 720
QY 721 AEXSSWVRPCLPASHVFPAGKAIWYTGNGHTOYGTGALLQKGIIRVYNQTTGENTL 780
D 721 VERSTVVRPCLPASHVFPAGKAIWYTGNGHTGEGGIALILQKGIIRVYNQTTGENTL 780
QY 781 PQOITPRMNCVGLSGVDSCQDSSGGLPSSVEADGRIFGAGVVSWMGDGAGRNKPVYTT 840
D 781 PQOITPRMNCVGLSGVDSCQDSSGGLPSSVEADGRIFGAGVVSWMGDGAGRNKPVYTT 840
QY 841 RLP 843
D 841 RLP 843
```

RESULT 5  
US-09-654-600A-10

Sequence 10, Application US/09654600A  
Patent No. 6649741  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
FILE REFERENCE: D6064CIP/D  
CURRENT APPLICATION NUMBER: US/09/654,600A  
PRIORITY FILING DATE: 2000-09-01  
PRIORITY FILING DATE: 09/02/2003  
PRIORITY FILING DATE: 1999-10-20  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 10  
LENGTH: 902  
TYPE: PRT  
ORGANISM: Mus musculus  
FEATURE:  
OTHER INFORMATION: Epithin  
US-09-654-600A-10

Query Match 81.0%; Score 3793; DB 4; Length 902;  
Best Local Similarity 81.1%; Pred. No. 4,86-269;  
Matches 684; Conservative 71; Mismatches 88; Indels 0; Gaps 0;

```
QY 1 MGSBRARKGKGGPKDGAAGLKYSRHRKVNGLSEGVFLPVNNVKYKXKGPGRWVLLAA 60
D 1 MGSBRARKGKGGKQDPAAGLKYSRLNNMNGFESEVFLPANNAKKXKXKGPGRWVLLVA 60
QY 61 VLIGLLVLIGLIGFLVWHLQYRDVRYOKVKNQYMRITNENFVDAYENSNSTEFVSLASKV 120
D 61 VLSEFLLSLMAGLLVWHFHYRNVRYOKVFNQHLRITNEIFLDAYENSTSTEFVSLASKV 120
QY 121 KDALKLLYSQVPLGPRYHKESAVTAESGVIAIYWSSEFSIPHLYBEAERWMAEERVVM 180
D 121 KEALKLLYNEVPVLGPRYHKESAVTAESGVIAIYWSSEFSIPHLYBEAERWMAEERVVM 180
QY 181 LPPRARSLSFVYTSVVAFPPTDSKTVOPTODNSCSFGLHARGVELMRFTTGPSPSPYPA 240
D 181 LPPRARSLSKSFVTSVVAFPIDPMLQRTODNSCSFALHARGAAYVTRFTTGPSPSPYPA 240
QY 241 HARCQWALRGDADSVLSLTFERSFDLASCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300
D 241 HARCQWALRGDADSVLSLTFERSFDVAPCDEHGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300
QY 301 YNLTFFSSQVNVLLITLITNTERRHHPGEATFFQLPRMSSCGGLRLKAQGTFFNSPYPGHY 360
D 301 YNLTFFLSQVNVFLVTLITNTERRHHPGEATFFQLPRMSSCGGLRLKAQGTFFNSPYPGHY 360
QY 361 PPNIDCTWNTNIEVPNNQHVKKRFEFYLLBPGVAGTCKPDYVEINERKXCGESRQFVYTS 420
D 361 PPNINCTWNTNIEVPNNRNVKRFKLFYLVDPNVVPGSLCKDYVEINERKXGESRQFVYTS 420
QY 421 NSNKTIVRFHSDSYTDTGFLAEYLSYSSDPGPGFTCTGRCIRKELRCDGMADCTDH 480
D 421 NSSKITVYHFHSDSYTDTGFLAEYLSYSSDPGPGFTCTGRCIRKELRCDGMADCTDH 480
QY 481 SDELINCSGADGHQFTCKNFKCKPLFWVCDVNDGSDNSDEQSCCPAQOTFRCSNGKCLSK 540
D 481 SDEYRCRCAVTHQFTCKNFKCKPLFWVCDVNDGSDNSDEQSCCPAGSFKCSNGKCLPQ 540
QY 541 SQQCKNGKDDGSDSDASCPKVVNVVCTKHTYRCLNGLCLSKNPGCDGKEDSDGSDSK 600
D 541 SQQCKNGKDDGSDSDASCDVNVVNSCTKHTYRCLNGLCLSKNPGCDGKEDSDGSDSK 600
QY 601 DCDGGLRSFTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPNMLVSAAHCYID 660
D 601 NCCGGLRSFTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPNMLVSAAHCFOD 660
QY 661 DRGRYSDPTQWTAFLGLHDSQSRAPGVQERRLKRIISHPFNDFTFYDIALLELEKP 720
D 661 DRKFKYSDYTMWTAFLGLHDSQSRASGVQELKRIIHPFNDFTFYDIALLELEKS 720
```

Db 661 DKNFYSQYTWMTAFLGLDQSKSASGVQELKRLITTHPSFNDFTFDYDIALLELEKS 720  
Qy 721 AEYSMWPRPCLPDASHVFPKAIWWTGHTQYGGTGAIILOKGEIRVINTTCENLL 780  
Db 721 VEYSTVVRPCLPDATHVFPKAIWWTGHTQYGGTGAIILOKGEIRVINTTCEDLM 780  
Qy 781 POQITPRMVCVGLSGVDSGCGSPSSVEADGRIFGAGVSWGDCGAKRNPQYTT 840  
Db 781 POQITPRMVCVGLSGVDSGCGSPSSVEADGRIFGAGVSWGDCGAKRNPQYTT 840  
Qy 841 RLP 843  
Db 841 RLP 843

RESULT 6  
US-09-657-986B-2  
Sequence 2, Application US/09657986B  
Patent No. 6797504  
GENERAL INFORMATION:  
APPLICANT: Madison, Edwin L.  
APPLICANT: Semple, Joseph Edward  
APPLICANT: Coombs, Gary Samuel  
APPLICANT: Reiner, John Eugene  
APPLICANT: Ong, Edgar O.  
APPLICANT: Araldi, Gian Luca  
TITLE OF INVENTION: Inhibitors of Serine Protease Activity of Matricriase or  
TITLE OF INVENTION: MTSP1  
FILE REFERENCE: Corvas 255/049  
CURRENT APPLICATION NUMBER: US/09/657, 986B  
CURRENT FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 2  
LENGTH: 241  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-09-657-986B-2

Query Match 27.9%; Score 1305; DB 4; Length 241;  
Best Local Similarity 99.2%; Pred. No. 9.2e-88;  
Matches 239; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 615 VVGSTDADEGEMPOVSLHALGQGHICGASLISPMWLVSAHCYIDDSGFRYSPTQWTA 674  
Db 1 VVGSTDADEGEMPOVSLHALGQGHICGASLISPMWLVSAHCYIDDSGFRYSPTQWTA 60  
Qy 675 FLGLHDSQSRAPGQVERRLKRIISHPPFNFTPDYDIALLELEKPAEYSSMWPRICLPD 734  
Db 61 FLGLHDSQSRAPGQVERRLKRIISHPPFNFTPDYDIALLELEKPAEYSSMWPRICLPD 120  
Qy 735 ASHVFPAGKAIWWTGHTQYGGTGAIILOKGEIRVINTTCENLLPOQITPRMVCVGL 794  
Db 121 ASHVFPAGKAIWWTGHTQYGGTGAIILOKGEIRVINTTCENLLPOQITPRMVCVGL 180  
Qy 795 SGGVDSGCGSPSSVEADGRIFGAGVSWGDCGAKRNPQYTTPLPRDWTKEKNTG 854  
Db 181 SGGVDSGCGSPSSVEADGRIFGAGVSWGDCGAKRNPQYTTPLPRDWTKEKNTG 240  
Qy 855 V 855  
Db 241 V 241

RESULT 7  
US-08-200-900A-2  
Sequence 2, Application US/08200900A  
Patent No. 5655566  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc. - Legal Affairs  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/200,900A  
FILING DATE: 23-FEB-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Melner, Maureen C.  
REGISTRATION NUMBER: 31,544  
REFERENCE/DOCKET NUMBER: GI 5201-FWC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 876-1170 X8574  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 798 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-200-900A-2

Query Match 15.0%; Score 700.5; DB 1; Length 798;  
Best Local Similarity 32.6%; Pred. No. 7.2e-43;  
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;

Qy 333 QLPKMSGCG--RLRKAQGTFFNSPYPGHYPPNIDCTWNTIEVPNNQYKARFKFYLLEP 390  
Db 298 ELF--TDGGPHDMEWNTFTSINFPNSVQNFQIMNLNAQKQIQLHFOGFDLENI 355  
Qy 391 GVPAGTCRKDYVEI--NBEKYCGERSQV-----VTSNKNKTYTRPHSQSTYDT 438  
Db 356 A-----DVEIRDE--GDSLFLAVYTGPPVNDVFTNRMVTFETDMLAKQ 404  
Qy 439 GLFLAEVLS--YDSSDC--PQOFTCRGRCIRKELRCDGMADCTDHSDELNC-----SC 488  
Db 405 GPKMFTTGGLGIPBCKEDNFQCKDGBCLPLVNLCDGFPHCKDSDBAHCALFNGIT 464  
Qy 489 DAGH--OFTCKNKECKELFWVCDVNDGDNSEDGSCCPAQTPRCNGKCLSKSQOQNG 546  
Db 465 DSSGLVQFRIS-----IMHV-----ACAEK-----WTTQISDDVC-----QLIG 499  
Qy 547 KDCGSDSDEASC-----PKVNVVCTKHTYRCLNGLCLSKNPECDKEDGSDSDEK 600  
Db 500 ---LGTNNSVPTPSTGGGPPVNTAP-----NSSLILTPSQOQ-----LEDSILL 544  
Qy 601 DCD---GGLSFTQA---RVVGTDADEGEMPOVSLHALGQGHICGASLISPMWLVSA 655  
Db 545 QCNKSKGKLVTVQESPKTVGSSDSREGAIPWVALVFPDQ--QVCASLVSRLMVSAA 603  
Qy 656 HCYIDRGRFYSPTQWTAFLGLHDSQSRAPGQVERRLKRIISHPPFNFTPDYDIAL 715  
Db 604 HCYVG-----RMERSKKAVALGLMAISNLISPOLIETLIQDIIVNPHNKRKRKNDLMM 659  
Qy 716 ELEKPAEYSSMWPRICLPDASHVFPKAIWWTGHTQYGGTGAIILOKGEIRVINTTC 775  
Db 660 HLEKRVVTVDTQICLPEENQVPPPGRICSIAGMGALIVGSTADVLQEADEVLLSEK 719  
Qy 776 CENLLPO--QITPRMVCVGLSGVDSGCGSPSSVEADGRIFGAGVSWGDCGAKRNP 834  
Db 720 CQOQMPENITENNVCAGYEAGVDSGCGSGGLMCOE--NNRMLLAGVTSFGYQCALPN 778  
Qy 835 KPGVYTRLPFRDWIX 850



Db 604 HCYVG----RAMEPSKMAVGLHMASSLTSPQIETRLIDQIVINPHYKKNNDIAMM 659  
Qy 716 ELEKPAEYSSMWRLCPDASHVFPAGKAIWYMGHTQYGTGALLIIOKEIRVINGTT 775  
Db 660 HLEKKNVTYDIOPICLPEENQVPPPGRICIAGGALITYGSDTADVIQADAVPLLSNEK 719  
Qy 776 CENLLPQ-QITPRMVCVGLSGVDSCQDGGPPLSSVEADGRIFGAGVSMGDCAGRN 834  
Db 720 CQOQMPENTENNVCAGYEAGVDSCQDGGPPLMCE-NNRWLLAGVTSFGYCALPN 778  
Qy 835 KPGVYTRLPFRDWIK 850  
Db 779 RGVYARVPFRTEWIO 794

## RESULT 10

US-09-734-675-4  
; Sequence 4, Application US/09734675  
; Patent No. 6365391  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; FILE REFERENCE: CLO00862  
; CURRENT APPLICATION NUMBER: US/09/734,675  
; NUMBER OF FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 407  
; TYPE: PRT  
; ORGANISM: Human  
US-09-734-675-4

Query Match 12.4%; Score 583; DB 3; Length 407;  
Best Local Similarity 45.9%; Pred. No. 1.2e-34;  
Matches 119; Conservative 37; Mismatches 83; Indels 20; Gaps 6;

Qy 604 CGLR--SFTROARVVGSTDADEGEWPMQVSLHAGCHICGASLISPMVLVAHACYID 660  
Db 162 CGTRRSKTLGSLRIVGGTEVEGEWPMQASLQMDG-SHRCGATLINATWLVSAHCP- 219  
Qy 661 DRGRYSDPTQWTAFLGLHDSQSRAPVOERLKRITSHFPNDFPTDYDIALLEK 720  
Db 220 ---TYKNPAAWTASFGV----TIKPSKMRGLRRIIVHEKXKPHSDYDISLAELSSP 270  
Qy 721 AEYSMWRLCPDASHVFPAGKAIWYMGHTQYGTGALLIIOKEIRVINGTTCEML 780  
Db 271 VPTNNAHVRLCPDASHVFPAGKAIWYMGHTQYGTGALLIIOKEIRVINGTTCEML 328  
Qy 781 PQ---QITPRMVCVGLSGVDSCQDGGPPLSSVEADGRIFGAGVSMGDCAGRN 836  
Db 329 PQAVNDATTPRLMCLGSLGKTDACQDGGPPLSSVDARDIWLGLAGVSMGDCAGRN 388  
Qy 837 GVTYTRLPFRDWIK 855  
Db 389 GVTYTRVTRALRDMITSKGI 407

## RESULT 11

US-09-685-166A-895  
; Sequence 895, Application US/09685166A  
; Patent No. 6630305  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jianshun  
; APPLICANT: Dillon, Devin C.  
; APPLICANT: Mitchem, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jieng, Yugu  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Patrick  
; APPLICANT: Li, Samuel  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yaser A.W.  
; APPLICANT: Hepler, William  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
; FILE REFERENCE: 210121.427C21  
; CURRENT APPLICATION NUMBER: US/09/685,166A  
; CURRENT FILING DATE: 2000-10-10  
; NUMBER OF SEQ ID NOS: 898  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 895  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-685-166A-895

Query Match 11.9%; Score 557.5; DB 4; Length 492;  
Best Local Similarity 33.9%; Pred. No. 1.1e-32;  
Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;

Qy 522 GCSGPAQTRFC-SNGKLSKSGCNGKDDCGDSDSEASCKRV--NIV----- 565  
Db 110 GSKCSNGIBCDSSGCTINPSMCDGVSHCPGGEDENRVRLVGPFLQMYSSQKSMH 169  
Qy 566 -----TCT---KATYRCANG-----LCSKGNPECDKEDCSD 595  
Db 170 PVCDDPMNENYGRACACDMGKKNFYSSQGIIVDSGTSFMKNTSGANVDIYKLYHSD 229  
Qy 596 GSDEK-----DCDGLRSFTRQARVVGSTDADEGEWPMQVSLHAGCHICGASLIS 647  
Db 230 ACSKAVSLRCLACGVNLNS-SRQSRIVGSESLPAMPWQVSLH-VQVHVCGGSLIT 287  
Qy 648 PNLVSAHACYIDRGRYSDPTQWTAFLGLHDS-QSRAPVOERLKRITSHFPNDF 706  
Db 288 PEMVITVAHCEK-----LNNPMHTVAFAGILRSFMYAGAYQ--VQKVISHPNYSK 340  
Qy 707 TFDVIALLELEKPAEYSSMWRLCPDASHVFPAGKAIWYMGHTQYGTGALLIIOK 766  
Db 341 TKANDILMLKQLTNDLVKPYCLPFGMMLQPELCHISGAGLEBKSEVLNAA 400  
Qy 767 EIRVINGTTCEML-LPQOITPRMVCVGLSGVDSCQDGGPPLSSVEADGRIFG-AGV 823  
Db 401 KVLIERFORCNSRYVNDLITPAMICAGFLQGVNDSQDGGPPL--VTSNNIMWLIGD 458  
Qy 824 VSMGDCAGRNKPGVYTRLPFRDWI 849  
Db 459 TSMGSCAKAYRPGVYGNVWVFTDWI 484

## RESULT 12

US-09-879-792-14  
; Sequence 14, Application US/09879792  
; Patent No. 6734006  
; GENERAL INFORMATION:  
; APPLICANT: Xiao, Yonghong  
; APPLICANT: Gedrich, Richard  
; TITLE OF INVENTION: Regulation of Human Transmembrane Sertine  
; TITLE OF INVENTION: Protease  
; FILE REFERENCE: 02973.00035  
; CURRENT APPLICATION NUMBER: US/09/879,792  
; CURRENT FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/211,224  
; PRIOR FILING DATE: 2000-06-13  
; PRIOR APPLICATION NUMBER: US 60/283,353  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: US 60/283,648  
; PRIOR FILING DATE: 2001-04-16



```

; (Docket No. 6734006 LIO-81-W0)
; PRIOR APPLICATION NUMBER: PCT
; PRIOR FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PaatSeq for Windows Version 4.0
; SEQ ID NO: 14
; LENGTH: 492
; TYPE: FRT
; ORGANISM: Homo sapiens
; IS-09-879-792-14

```

Query Match	11.9%	Score 557.5;	DB 4;	Length 492;
Best Local Similarity	33.9%	Pred. No. 1,1e-32;		
Matches 131;	Conservative 57;	Mismatches 129;	Indels 69;	Gaps 14
QY	522	GCSCPAQTRC-SNGKCLSKSQOQCKDDCGDSDPAACPKV--NVV-----	565	
Db	110	GSKCSNSGIEDCSGCTCINPSCMGVSHCPGEGEDENRCVRLYGFNFILQMYTSQPKSMH	169	
QY	566	-----TCT-----KHTYRLCNG-----LCLSKGPEDEGKEDCD	595	
Db	170	PVCODDMNENYGRAACRDMGYKNNYSSQGIYDSDGSTSFMKLTNSAGVNDLYKKLYHSD	229	
QY	596	GSDCK-----DCCGLRSFTROARVVGTDADGEMPMQVSLHALQGGHICGASLIS	647	
Db	230	ACSSKAVYSLRCLACGVNLNS-SRGSRLVGGESALPGAMPQVSLH-VQNVAVCGSSILT	287	
QY	648	PNMLVSAHCTIYDRGFRYSIPTQMTALFLGHDG-QSAPVQGERLKRITISHPFNDF	706	
Db	288	PEWITVAHACEVK---LNNPMWHTAPAGILRQSFMYGAGYQ---YQKVLISHPYDSK	340	
QY	707	TFEDVIALLEVEKPEAYSSMYRPLCLPASHVFPAGKAIWVGWGHGTGYGAGALILQKG	766	
Db	341	TKNNDIALMKIQKPLTFNDLYKPVCLTPMPGMMLQGEOLCWMISGWATEBKGTSEVLNAA	400	
QY	767	EIRVINOQTGEN--LLPQOITPRMWCVFLSGVDSCGSDSGPLSSYEADRIFG-AGV	823	
Db	401	KVLLIETQRCSRYVDMLITPAMICAFPLQGNVSDCGDSGGLP--VTSNNNIWMLIGD	458	
QY	824	VSMGDGCGAKRKPGVYTRTLPLFRMDI	849	
Db	459	TSWGSQCAKAYRPGVYGNVWVFTDII	484	

```

RESULT 13
US-09-679-426-895
Sequence 895, Application US/09679426
Patent No. 6759515
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427620
CURRENT APPLICATION NUMBER: US/09/679,426
CURRENT FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 895
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 895

```

```

; LENGTH: 492
;
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-679-426--895

```

[illegible]

```

US-09-342-749-2
US-09-342-749-2
Sequence 2, Application US/09342749
Patent No. 6166194
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TP53RS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-342-749-2

Query Match          11.9%; Score 555.5; DB 3; Length 492;
Beet Local Similarity 33.9%; Pred. No. 1.6e-32;
Matches 131; Conservative 56; Mismatches 110; Indels 69; Gaps 14;

QY      522  GCSCPAQTFRC-SNGKCLSKSQGCGKDDCGDGSDEASCPKV---NVV----- 565
          :::::
Db      110  GSKCSNGSIGECDDSSGCTINPSNMWCDGVSHCPGGBEDNRCVRLYGPNFILQYSSORSKSWH 169
          :::::

QY      566  -----TCF-----KHTYRLCLNG-----LCTSKGNPCDGDKEDCSD 595
          :::::
Db      170  PVCDDMNENYGRACACDMGYKKNFYSSQGIVDSSGTSFMKLTNTSAGNVDIYKKLTHSD 229
          :::::

```



Db 116 ENQVFPGRNCISAGMGTVVYQGTANILQADVPLLSNERCQOQMBEYNTENMCAGY 175

QY 794 LSGGVDSGCGDSGGPILSSVEADGRIFGAGVYSWGDGAGRNKPGVYTRLPPLFRDWIK 850

Db 176 EAGGVDSGCGDSGGPILMCQE-NNRWFLAGVTSFGYQCALPNRPGVYARVPFTFEMIQ 231

RESULT 17

US-08-807-151-3

; Sequence 3, Application US/08807151

; Patent No. 6043033

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Lal, Preeti

; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED

; TITLE OF INVENTION: PROTEASE

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: US

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/807.151

; FILING DATE: Filed Herewith

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0227 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 235 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 416132

US-08-807-151-3

Query Match 11.5%; Score 540.5; DB 3; Length 235;

Best Local Similarity 43.5%; Pred. No. 7.6e-32;

Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

QY 615 VVGSTDADEGEWMPQVSLHALGQHICGASLISPNMLVSAHCVYIDDRGFRYSDFPTQWTA 674

Db 1 IVGGSDBREGAMWVVALYFDQ-QVCGASLVSRLDWLVSAAHCYV---RNNEPSKMXA 55

QY 675 FLAGHDSQSAFQVQERRLRKRIISHPFNDFTFDYDIALLEKPAEYSSMWVPICLPD 734

Db 56 VLGHHMASNLTSQIEIRLIDQIVINPHYKRRKNDIAMHLEMKVNTDYIOPICLPE 115

QY 735 ASHVFPAGKAIWYTGWGHTOYGGTGALLQKEIRVINQTTCENLDP-QITPRMVCVF 793

Db 116 ENQVFPGRNCISAGMGTVVYQGTANILQADVPLLSNERCQOQMBEYNTENMCAGY 175

QY 794 LSGGVDSGCGDSGGPILSSVEADGRIFGAGVYSWGDGAGRNKPGVYTRLPPLFRDWIK 850

Db 176 EAGGVDSGCGDSGGPILMCQE-NNRWFLAGVTSFGYQCALPNRPGVYARVPFTFEMIQ 231

RESULT 18

US-09-478-957-3

; Sequence 3, Application US/09478957

; Patent No. 6350448

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Lal, Preeti

; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED

; TITLE OF INVENTION: PROTEASE

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: US

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/478,957

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/807,151

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0227 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 235 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 416132

US-09-478-957-3

Query Match 11.5%; Score 540.5; DB 3; Length 235;

Best Local Similarity 43.5%; Pred. No. 7.6e-32;

Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

QY 615 VVGSTDADEGEWMPQVSLHALGQHICGASLISPNMLVSAHCVYIDDRGFRYSDFPTQWTA 674

Db 1 IVGGSDBREGAMWVVALYFDQ-QVCGASLVSRLDWLVSAAHCYV---RNNEPSKMXA 55

QY 675 FLAGHDSQSAFQVQERRLRKRIISHPFNDFTFDYDIALLEKPAEYSSMWVPICLPD 734

Db 56 VLGHHMASNLTSQIEIRLIDQIVINPHYKRRKNDIAMHLEMKVNTDYIOPICLPE 115

QY 735 ASHVFPAGKAIWYTGWGHTOYGGTGALLQKEIRVINQTTCENLDP-QITPRMVCVF 793

Db 116 ENQVFPGRNCISAGMGTVVYQGTANILQADVPLLSNERCQOQMBEYNTENMCAGY 175

QY 794 LSGGVDSGCGDSGGPILSSVEADGRIFGAGVYSWGDGAGRNKPGVYTRLPPLFRDWIK 850

Db 176 EAGGVDSGCGDSGGPILMCQE-NNRWFLAGVTSFGYQCALPNRPGVYARVPFTFEMIQ 231

RESULT 19

US-09-518-046-2

; Sequence 2, Application US/09518046





Db 6 SAGVNDIYKLYHSDAGSSKAVLSRCLACGVNLNS-SRQSRIVGSEBALGAMPVQSL 64  
Qy 633 HALGQGHICGASLISPMVLVSAHCYIDDRGFRYSPTQWTAFLGLHDQS-QRSAPGYOE 691  
Db 65 H-VQNVHVCSSGSIITPEIYVTAACVKEKP-----LNNPMHMTAFAGILQSFMFYGAGYQ- 118  
Qy 692 RLRLRIISHPPNFPTDYDIALLELEKPAVSSMVRPICKPDASHVPAKAIWYQWG 751  
Db 119 --VEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKVCPLPNQPMMLQPEOLCMISGMG 176  
Qy 752 HTQYGGGALLIQKEIRIVNOTTCEN--LLPQOITPRMVCVFLSGVDSGQSDGGFL 809  
Db 177 ATEKRGKTSBYLNAKVLITFORCNSRYVDNLITPMICAGFLQGVDSGQSDGGXL 236  
Qy 810 SSVSEADGRIFG-AGVSVWGDCAGRNKPGVYTRLLPLFRDWT 849  
Db 237 --VTSKNIMWMLIGDTSWSSGCAKAYRPGVYGNVWVFTDWI 275

## RESULT 24

US-10-177-661-2  
Sequence 2, Application US/10177661  
Patent No. 6794173  
GENERAL INFORMATION:  
APPLICANT: Anderson, Dirk M.  
TITLE OF INVENTION: DENDRITIC CELL TRANSMEMBRANE SERINE PROTEASE  
FILE REFERENCE: 3256-A  
CURRENT APPLICATION NUMBER: US/10/177,661  
CURRENT FILING DATE: 2002-06-20  
PRIOR APPLICATION NUMBER: US 60/299,606  
PRIOR FILING DATE: 2001-06-20  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 477  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-177-661-2

Query Match 10.9%; Score 509.5; DB 4; Length 477;  
Best Local Similarity 32.3%; Pred. No. 3.5e-29;  
Matches 134; Conservative 50; Mismatches 133; Indels 93; Gaps 17;

Qy 464 CIRKELRCGWDCTDHSDELNC-----SCDAGHQTCKNFKCKPLFW--VCD 509  
Db 114 CPKHAVRCDGVDDCKLSDELGCVRPMDKSLIKYSGSSHQ-----WLPICS 161  
Qy 510 SVNDGNSDQGGSCPAQTRCNSGKCLSKSQCGNKDQGGSDASCPKRVNVYCTK 569  
Db 162 S--WNDSYSSEKTC-----QQLGFSAHRTTEVHRDPANFSILR 200  
Qy 570 HTYRCLNGLCLSKGNPECDGKE---DCSDGSEKDCDCGLRSFTROARVVGTDADGE 625  
Db 201 YNSTIQLSLHRS---ECPSQRYISLQCS-----HCGLRAMT--GRIVGALLASDSK 246  
Qy 626 WPMQVSLH-FCGTHICGGLTLDAQWVLTAAHCFVTRKYLEG-----WKVYAGTSN 297  
Db 247 WPMQVSLH-FCGTHICGGLTLDAQWVLTAAHCFVTRKYLEG-----WKVYAGTSN 297  
Qy 678 LHDQSQSAAGVQERRLKRIISHPPNFPTDYDIALLELEKPAVSSMVRPICKLPDASH 737  
Db 298 LHQPLEAS-----IAEIIINSNYTDEBDYDIALMRLSKPLTSLAHHPACLPWNGQ 350  
Qy 738 VFPAKAIWYMGHT-QYGGTGALLIQKEIRIVNOTTCENLL--PQOITPRMVCVFL 794  
Db 351 TFSINETCWITGFKETFTDXTSPFLREVOVNLIDPKKNDYLVYDSYILPRMCGADL 410  
Qy 795 SGGVDSGQSDGGFLSSVEADGRIFGAGVSVWGDCAGRNKPGVYTRLLPLFRDWT 849  
Db 411 RGRDSCGSDGGFL--VCEQNNRWYLAGVTSWGTGCGGRNKPQVYTKVTEVLPMI 464

## RESULT 25

US-09-879-792-12  
Sequence 12, Application US/09879792  
Patent No. 6734006  
GENERAL INFORMATION:  
APPLICANT: Xiao, Yonghong  
TITLE OF INVENTION: Regulation of Human Transmembrane Serine  
PROTEASE  
FILE REFERENCE: 02973.00035  
CURRENT APPLICATION NUMBER: US/09/879,792  
CURRENT FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/211,224  
PRIOR FILING DATE: 2000-06-13  
PRIOR APPLICATION NUMBER: US 60/283,353  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: US 60/283,648  
PRIOR FILING DATE: 2001-04-16  
PRIOR APPLICATION NUMBER: PCT  
PRIOR FILING DATE: 2001-06-12  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 562  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-879-792-12

Query Match 10.9%; Score 509.5; DB 4; Length 562;  
Best Local Similarity 32.3%; Pred. No. 4.3e-29;  
Matches 134; Conservative 50; Mismatches 138; Indels 93; Gaps 17;

Qy 464 CIRKELRCGWDCTDHSDELNC-----SCDAGHQTCKNFKCKPLFW--VCD 509  
Db 199 CPKHAVRCDGVDDCKLSDELGCVRPMDKSLIKYSGSSHQ-----WLPICS 246  
Qy 510 SVNDGNSDQGGSCPAQTRCNSGKCLSKSQCGNKDQGGSDASCPKRVNVYCTK 569  
Db 247 S--WNDSYSSEKTC-----QQLGFSAHRTTEVHRDPANFSILR 285  
Qy 626 WPMQVSLH-FCGTHICGGLTLDAQWVLTAAHCFVTRKYLEG-----WKVYAGTSN 382  
Db 332 WPMQVSLH-FCGTHICGGLTLDAQWVLTAAHCFVTRKYLEG-----WKVYAGTSN 382  
Qy 678 LHDQSQSAAGVQERRLKRIISHPPNFPTDYDIALLELEKPAVSSMVRPICKLPDASH 737  
Db 383 LHQPLEAS-----IAEIIINSNYTDEBDYDIALMRLSKPLTSLAHHPACLPWNGQ 435  
Qy 738 VFPAKAIWYMGHT-QYGGTGALLIQKEIRIVNOTTCENLL--PQOITPRMVCVFL 794  
Db 436 TFSINETCWITGFKETFTDXTSPFLREVOVNLIDPKKNDYLVYDSYILPRMCGADL 495  
Qy 795 SGGVDSGQSDGGFLSSVEADGRIFGAGVSVWGDCAGRNKPGVYTRLLPLFRDWT 849  
Db 496 RGRDSCGSDGGFL--VCEQNNRWYLAGVTSWGTGCGGRNKPQVYTKVTEVLPMI 549

## RESULT 26

US-08-508-448C-25  
Sequence 25, Application US/08508448C  
Patent No. 5804410  
GENERAL INFORMATION:  
APPLICANT: Kazuyoshi YAMAKA et al.  
TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING  
TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack







**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:14:53 ; Search time 122.858 Seconds  
(without alignments)  
2496.496 Million cell updates/sec

Title: US-09-936-333-27

Perfect score: 4683  
Sequence: 1 MSGDRAKKGCGGPKRFRAGL.....PGVYTRLPFRDWIKENTGV 855

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues  
Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database :

1: Geneseqp\_23Sep04:\*  
2: geneseqp1980s:\*  
3: geneseqp1990s:\*  
4: geneseqp2000s:\*  
5: geneseqp2001s:\*  
6: geneseqp2002s:\*  
7: geneseqp2003as:\*  
8: geneseqp2003bs:\*  
9: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4659	99.5	855	3 AAB19552	Aab19552 Human mat
2	4659	99.5	855	4 AAB35465	Aab35465 Human mat
3	4659	99.5	855	5 ADI16817	Adi16817 Human NOV
4	4659	99.5	855	5 ADI16883	Adi16883 Human NOV
5	4659	99.5	855	5 ADI16876	Adi16876 Human NOV
6	4659	99.5	855	7 ADN19867	Adn19867 Cancer/an
7	4659	99.5	855	8 ADN04754	Adn04754 Antiproti
8	4655	99.4	855	5 ADI16884	Adi16884 Human NOV
9	4655	99.4	855	5 ADI16818	Adi16818 Human NOV
10	4653	99.4	855	2 AAY06671	Aay06671 Tumour an
11	4653	99.4	855	4 AAB98500	Aab98500 Human TAD
12	4653	99.4	855	5 AAE06930	Aae06930 Human mem
13	4653	99.4	855	5 AAO23929	Aao23929 Type II t
14	4653	99.4	855	5 ADI16816	Adi16816 Human NOV
15	4653	99.4	855	5 ADI16882	Adi16882 Human NOV
16	4653	99.4	855	5 ADI16875	Adi16875 Human NOV
17	4653	99.4	855	6 AAB56619	Aab56619 Human mem
18	4653	99.4	855	6 AAO30146	Aao30146 Human mem
19	4653	99.4	855	6 AAE29820	Aae29820 Human mem
20	4653	99.4	855	6 AAE29791	Aae29791 Human mem
21	4653	99.4	855	6 ABP72376	Abp72376 Transmemb
22	4653	99.4	855	7 ADB97551	Adb97551 Human MTS
23	4653	99.4	855	7 ADI10371	Adi10371 Human cel
24	4653	99.4	855	8 ADG65326	Adg65326 Human MTS
25	4653	99.4	855	8 ADI28861	Adi28861 Human mat

26	4653	99.4	855	8 ADJ46895	Adj46895 Human tra
27	4625	98.8	853	8 ADO55145	Ado55145 Protein #
28	4614	98.5	851	4 AAM25628	Aam25628 Human pro
29	4614	98.5	851	4 ABB11428	Abb11428 Human mem
30	4302	91.9	932	4 ABG21442	Abg21442 Novel hum
31	4167	89.0	762	3 AAY90284	Aay90284 Human pep
32	4158.5	88.8	782	5 ABG96427	Abg96427 Human ova
33	4092	87.4	757	5 ADI16508	Adi16508 Human NOV
34	4057	86.6	757	8 ADM42162	Adm42162 Human nov
35	3884	82.9	855	5 AAE23083	Aae23083 Epithin p
36	3884	82.9	855	5 ADI16819	Adi16819 Murine NO
37	3864	82.6	855	5 ADI16877	Adi16877 Murine NO
38	3866	82.6	855	5 ADI16820	Adi16820 Rat NOVX
39	3866	82.6	855	5 ADI16881	Adi16881 Rat NOVX
40	3866	82.6	855	5 ADI16878	Adi16878 Rat NOVX
41	3793	81.0	902	4 AAB38507	Aab38507 Murine ep
42	3793	81.0	902	5 AAU80517	Aau80517 Mouse epi
43	3793	81.0	902	5 AAU77549	Aau77549 Murine ty
44	3773	80.6	633	3 AAB19551	Aab19551 Human mat
45	2986	63.8	620	3 AAB37748	Aab37748 Human can
46	2660	56.8	845	5 ADI16879	Adi16879 African c
47	1343	28.7	362	4 ABG21441	Abg21441 Novel hum
48	1305	27.9	241	4 AAE06936	Aae06936 Human mem
49	1305	27.9	241	5 AAO22930	Aao22930 Type II t
50	1305	27.9	241	5 AAE22837	Aae22837 Human mat
51	1305	27.9	241	6 ABP56620	Abp56620 Human MTS
52	1305	27.9	241	6 AAO30147	Aao30147 Human mem
53	1305	27.9	241	6 AAE29821	Aae29821 Human mem
54	1305	27.9	241	6 AAE29792	Aae29792 Human mem
55	1305	27.9	241	6 ABP72377	Abp72377 Transmemb
56	1305	27.9	241	7 ADA25176	Ada25176 Human MTS
57	1305	27.9	241	7 ADB97553	Adb97553 Human MTS
58	1305	27.9	241	8 ADI28862	Adi28862 Human mat
59	1301	27.8	241	8 AAM22987	Aam22987 Human ser
60	117.5	23.9	795	5 ABP43952	Abp43952 Human PRO
61	114.5	23.8	799	5 ADI16874	Adi16874 Murine NO
62	114.5	23.8	799	5 ADI16880	Adi16880 Murine NO
63	1112	23.7	802	3 AAY41710	Aay41710 Human PRO
64	1112	23.7	802	3 AAB44265	Aab44265 Human PRO
65	1112	23.7	802	3 AAB24052	Aab24052 Human PRO
66	1112	23.7	802	5 AAU82755	Aau82755 Amlno aci
67	1112	23.7	802	6 ABO25212	Abo25212 Novel hum
68	1112	23.7	802	6 ABU72218	Abu72218 Novel hum
69	1112	23.7	802	6 ABU84898	Abu84898 Human sec
70	1112	23.7	802	6 ABU61096	Abu61096 Human PRO
71	1112	23.7	802	6 ABU80365	Abu80365 Human sec
72	1112	23.7	802	6 ADA24708	Ada24708 Novel hum
73	1112	23.7	802	6 ABO19667	Abo19667 Novel hum
74	1112	23.7	802	6 ABA12369	Aba12369 Human sec
75	1112	23.7	802	6 ABO19558	Abo19558 Novel hum
76	1112	23.7	802	7 ADB73675	Adb73675 Human PRO
77	1112	23.7	802	7 ADB76391	Adb76391 Human PRO
78	1112	23.7	802	7 ADC43817	Adc43817 Human sec
79	1112	23.7	802	7 ADC61577	Adc61577 Human sec
80	1112	23.7	802	7 ADC63541	Adc63541 Human sec
81	1112	23.7	802	7 ADC66641	Adc66641 Human sec
82	1112	23.7	802	7 ADC68765	Adc68765 Human sec
83	1112	23.7	802	7 ADC68825	Adc68825 Human sec
84	1112	23.7	802	7 ADC67890	Adc67890 Human sec
85	1112	23.7	802	7 ADC41210	Adc41210 Human sec
86	1112	23.7	802	7 ADC67265	Adc67265 Human sec
87	1112	23.7	802	7 ADC62201	Adc62201 Human sec
88	1112	23.7	802	7 ADC41834	Adc41834 Human sec
89	1112	23.7	802	7 ADB49203	Adb49203 Human sec
90	1112	23.7	802	7 ADE35257	Ade35257 Human sec
91	1112	23.7	802	7 ADE16371	Ade16371 Human sec
92	1112	23.7	802	7 ADD72986	Add72986 Human sec
93	1112	23.7	802	7 ADD72344	Add72344 Human sec
94	1112	23.7	802	7 ADE16995	Ade16995 Human sec
95	1112	23.7	802	7 ADF47009	Adf47009 Human sec
96	1112	23.7	802	7 ADG52766	Adg52766 Human sec
97	1112	23.7	802	7 ADG60086	Adg60086 Human sec
98	1112	23.7	802	7 ADI60846	Adi60846 Human sec

99	1112	23.7	802	8	ADe48503	Human sec
100	1112	23.7	802	8	ADe89604	Human sec
ALIGNMENTS						
RESULT 1						
ID	AA19552 standard; protein; 855 AA.					
XX	AA19552;					
AC	AA19552;					
XX	22-JAN-2001 (first entry)					
XX	Human matrixase.					
XX	Human matrixase.					
XX	Matrixase; serine protease; human; breast cancer; pre-malignancy;					
KW	actinic keratosis; leukoplakia; Barrett's epithelium;					
KW	columnar metaplasia; ulcerative colitis; Bowenoid papulosis;					
KW	adenomatous colorectal polyp; Ovarian erythroplasia;					
KW	vulvar intraepithelial neoplasia; tumour; metastasis; therapy.					
XX						
OS	Homo sapiens.					
XX						
XX	Key					
XX	Location/Qualifiers					
FT	Region					
FT	214..327					
FT	/note="complement subcomponent 1r and 1s (C1r/s) region"					
FT	Peptide					
FT	249..251					
FT	/note="RCD motif"					
FT	Modified-site					
FT	302					
FT	/note="N-glycosylated"					
FT	Region					
FT	340..440					
FT	/note="complement subcomponent 1r and 1s (C1r/s) region"					
FT	Region					
FT	452..486					
FT	/note="LDL-receptor type region"					
FT	485					
FT	/note="N-glycosylated"					
FT	Region					
FT	487..523					
FT	/note="LDL-receptor type region"					
FT	Region					
FT	524..559					
FT	/note="LDL-receptor type region"					
FT	Region					
FT	566..602					
FT	/note="LDL-receptor type region"					
FT	Domain					
FT	603..855					
FT	/label=Catalytic_domain					
FT	Cleavage-site					
FT	614..618					
FT	/note="conserved proteolytic activation site"					
XX						
XX	WO200053232-A1.					
XX						
XX	14-SEP-2000.					
XX						
XX	10-MAR-2000; 2000MO-US006111.					
XX						
XX	12-MAR-1999; 99US-0124006P.					
XX						
XX	(GEOU ) UNIV GEORGETOWN.					
XX						
XX	Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;					
XX	P1					
XX	WPI; 2000-594268/56.					
XX						
XX	N-PSDB; AAA88493.					
XX						
XX	Treating malignancies, premalignant and pathologic conditions in a					
XX	PT					
XX	subject, comprises administering matrixase modulating agent.					
XX						
XX	Claim 14; Fig 16; 116pp; English.					
XX						
XX	The present sequence is that of the full-length form of human matrixase					
CC	a trypsin-like protease. This has an additional 172 amino acids compared					
CC	with the truncated form of matrixase given in AA19551. Either form can					
CC	be produced in transformed or transfected cells using a claimed method.					

CC	The zymogen (inactive) form of matrixase is a single-chain protein. The					
CC	active 2-chain form strongly interacts with fragments of a Kunitz-type					
CC	serine protease inhibitor (hepatocyte growth factor activator inhibitor,					
CC	HAI-1) to form SDS-stable complexes. In breast cancer cells, matrixase					
CC	is present mainly as the uncomplexed form. Only the complexed matrixase					
CC	in detected in human milk. The invention is directed to a method of					
CC	detecting a malignancy or a pre-malignant lesion in breast or other					
CC	tissue by detecting the presence of single- or 2-chain forms of					
CC	matrixase in the tissue. The object is to inhibit tumour onset, tumour					
CC	growth and metastasis. Malignancies and pre-malignant conditions					
CC	characterised by expression of the zymogen or activated form of					
CC	matrixase are treated by administering an inhibitor of matrixase,					
CC	especially a Bowman-Birk inhibitor. The pre-malignant condition is					
CC	atypical ductal hyperplasia of the breast, actinic keratosis,					
CC	leukoplakia, Barrett's epithelium of the esophagus, ulcerative colitis,					
CC	adenomatous colorectal polyps, erythroplasia of the Queyrat, Bowen's					
CC	disease, Bowenoid papulosis, vulvar intraepithelial neoplasia or					
CC	dysplastic changes to the cervix. The invention also provides methods for					
CC	in vivo or in vitro diagnosis of malignancy or pre-malignant lesion, and					
CC	methods of identifying matrixase modulators, including activators and					
XX	inhibitors					
XX						
SQ	Sequence 855 AA:					
	Query Match	99.5%;	Score 4659;	DB 3;	Length 855;	
	Best Local Similarity	99.6%;	Pred. No. 9,4e-308;			
	Matches 852; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;	
QY	1	MSGDRARKGGGKPKDGAAGIKYNSRHKYNGLEGGVPLPVNNVKKYKRGPRWVLLAA	60			
DB	1	MSGDRARKGGGKPKDGAAGIKYNSRHKYNGLEGGVPLPVNNVKKYKRGPRWVLLAA	60			
QY	61	VLIGLLVLLIGLIGFLVWHLQYRDVRYQKYNQYMTITNENFVDAYENNSTEVSLSKY	120			
DB	61	VLIGLLVLLIGLIGFLVWHLQYRDVRYQKYNQYMTITNENFVDAYENNSTEVSLSKY	120			
QY	121	KDALKLLYSGVPLPGYHKESAVTAPSEGSVVIYNSERSIPQHLVBEARVAAERVVM	180			
DB	121	KDALKLLYSGVPLPGYHKESAVTAPSEGSVVIYNSERSIPQHLVBEARVAAERVVM	180			
QY	181	LPPRARSLSFVTVSVVAPPTDSKTQVORTQDSCSGHLARGVELNRFPTPGPDSPPYA	240			
DB	181	LPPRARSLSFVTVSVVAPPTDSKTQVORTQDSCSGHLARGVELNRFPTPGPDSPPYA	240			
QY	241	HARQWALRGDADSVSLTFRSPDLASCDERGSDLTVVNTLSPMEPHALVQCGTYPPS	300			
DB	241	HARQWALRGDADSVSLTFRSPDLASCDERGSDLTVVNTLSPMEPHALVQCGTYPPS	300			
QY	301	YNLTFHSSQNVLLITLITNTERHHPGFATFPLPMSSCGRLRAQGFNPPYPGHY	360			
DB	301	YNLTFHSSQNVLLITLITNTERHHPGFATFPLPMSSCGRLRAQGFNPPYPGHY	360			
QY	361	PNVIDCTWNIENVNNOHVKKRFFYLLEBPGVAGTCPODYVEINGEKYCGESSQFVVS	420			
DB	361	PNVIDCTWNIENVNNOHVKKRFFYLLEBPGVAGTCPODYVEINGEKYCGESSQFVVS	420			
QY	421	NSNKITVRFSHSDSYDTGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCDGMADCTDH	480			
DB	421	NSNKITVRFSHSDSYDTGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCDGMADCTDH	480			
QY	481	SDELNCSGDAHQFTCKNFKCKPLFWVCDSVNDCCGNSDBQSCCAQTFRCGNGKLSK	540			
DB	481	SDELNCSGDAHQFTCKNFKCKPLFWVCDSVNDCCGNSDBQSCCAQTFRCGNGKLSK	540			
QY	541	SQOCNGKDDCGSDASCPKVNVVCTGHTYRCLNGLCLSKNPNPCDGEDSDSDSDK	600			
DB	541	SQOCNGKDDCGSDASCPKVNVVCTGHTYRCLNGLCLSKNPNPCDGEDSDSDSDK	600			
QY	601	DCDCGLRSFTQARVVGTDADGEMPMQVSLHALQGHICGASLISPMVLVSAACYYID	660			
DB	601	DCDCGLRSFTQARVVGTDADGEMPMQVSLHALQGHICGASLISPMVLVSAACYYID	660			
QY	661	DRCFRYSDPQTQWTFALGLHDQSORSAPOVERLKRILISHPPNDFTFYDIALLEKRP	720			

```

Db 661 DRGRYSDPQTQWTFGLGHDQSRAPGVQERRLKRIISHPFNDFEFDYDIALLELEKP 720
Qy 721 AEYSMWPRICLPDASHVFPAGKAIWYTGWGHYOGGTGALLIQKGIIRVYNQTCENLL 780
Db 721 AEYSMWPRICLPDASHVFPAGKAIWYTGWGHYOGGTGALLIQKGIIRVYNQTCENLL 780
Qy 781 PQQITPRMNCVGLSGVDSCQDSCGGLSSVEADGRIFGAGVYVWGDGAGRNKPGVYT 840
Db 781 PQQITPRMNCVGLSGVDSCQDSCGGLSSVEADGRIFGAGVYVWGDGAGRNKPGVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 2
AAB35465
ID AAB35465 standard; protein; 855 AA.
AC AAB35465;
XX
XX 06-JUN-2001 (first entry)
XX
XX 06-JUN-2001 (first entry)
XX
XX Human membrane-type serine protease MT-SPI.
XX
XX Human; membrane-type serine protease; MT-SPI; cancer.
XX
XX Homo sapiens.
XX
XX WO200123524-A2.
XX
XX PD 05-APR-2001.
XX
XX PF 02-OCT-2000; 2000MO-USO27250.
XX
XX PR 30-SEP-1999; 99US-00410362.
XX
XX PA (REGC ) UNIV CALIFORNIA.
XX
XX PI Craik CS, Takeuchi T, Shuman M;
XX
XX DR WPI; 2001-245002/25.
XX
XX PT N-PSDB; AAF28099.
XX
XX PT New nucleic acid encoding a membrane type serine protease, useful for the
XX
XX PT diagnosis, prognosis and treatment of cancer, particularly metastatic
XX
XX PS cancers.
XX
XX PS Claim 7; Fig 1; 102pp; English.
XX
XX CC The present invention provides the protein and coding sequences for the
XX
XX CC novel human membrane-type serine protease MT-SPI. Increased expression of
XX
XX CC this protein is associated with cancer, and so the sequences can be used
XX
XX CC in cancer diagnosis and the identification of treatments. The present
XX
XX CC sequence is the MT-SPI protein
XX
XX SQ Sequence 855 AA;

Query Match 99.5%; Score 4659; DB 4; Length 855;
Best Local Similarity 99.6%; Pred. No. 9,4e-308;
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 121 KDALKLYSGVPLGIPYHKESAVTAFSEGSVIAYWSEFSI PQHLYEAEERVAEERVVM 180
Qy 181 LPPRARSLSKPVVTSVVAFPPTDSKTQVORTDNSSCFGLHARGVLAFTTGPFPDPPYA 240
Db 181 LPPRARSLSKPVVTSVVAFPPTDSKTQVORTDNSSCFGLHARGVLAFTTGPFPDPPYA 240
Qy 241 HARCQMLARDADSVLSLTFERSFDLASCDERGSIDLVTYNTLSMEPHALVOLCGTYP 300
Db 241 HARCQMLARDADSVLSLTFERSFDLASCDERGSIDLVTYNTLSMEPHALVOLCGTYP 300
Qy 301 YNLTFFSSQVNLITLITNTERRHPGEATFPQLPRMSCGGRLRKAQGTFFNSPYTGHY 360
Db 301 YNLTFFSSQVNLITLITNTERRHPGEATFPQLPRMSCGGRLRKAQGTFFNSPYTGHY 360
Qy 361 PPNIDCTWNLTPVNNQHVKVPKFFYLEBGPVAGTCPKDYVEINSGKYGESQFVYTS 420
Db 361 PPNIDCTWNLTPVNNQHVKVPKFFYLEBGPVAGTCPKDYVEINSGKYGESQFVYTS 420
Qy 421 NSNKITVRPHSDSYTDTGFLAEYLSYSSSDPCPGQFTCTGRCIRKELRCDGADCTDH 480
Db 421 NSNKITVRPHSDSYTDTGFLAEYLSYSSSDPCPGQFTCTGRCIRKELRCDGADCTDH 480
Qy 481 SDEINCSGDAHQFTCKNKECKPLFWVCDSDVNDGSDSDGSCSPAQTRFCSNGKCLSK 540
Db 481 SDEINCSGDAHQFTCKNKECKPLFWVCDSDVNDGSDSDGSCSPAQTRFCSNGKCLSK 540
Qy 541 SQQNGKDDCGDSDSCPRVNVVCTKHTYRCINLCLSKNPECDGKEDSDGSDSK 600
Db 541 SQQNGKDDCGDSDSCPRVNVVCTKHTYRCINLCLSKNPECDGKEDSDGSDSK 600
Qy 601 DCDGGLRSFTRQARVVGTDADDEGEMPMQVSLHALGQGHICGASLISPNMLVSAHICYID 660
Db 601 DCDGGLRSFTRQARVVGTDADDEGEMPMQVSLHALGQGHICGASLISPNMLVSAHICYID 660
Qy 661 DRGRYSDPQTQWTFGLGHDQSRAPGVQERRLKRIISHPFNDFEFDYDIALLELEKP 720
Db 661 DRGRYSDPQTQWTFGLGHDQSRAPGVQERRLKRIISHPFNDFEFDYDIALLELEKP 720
Qy 721 AEYSMWPRICLPDASHVFPAGKAIWYTGWGHYOGGTGALLIQKGIIRVYNQTCENLL 780
Db 721 AEYSMWPRICLPDASHVFPAGKAIWYTGWGHYOGGTGALLIQKGIIRVYNQTCENLL 780
Qy 781 PQQITPRMNCVGLSGVDSCQDSCGGLSSVEADGRIFGAGVYVWGDGAGRNKPGVYT 840
Db 781 PQQITPRMNCVGLSGVDSCQDSCGGLSSVEADGRIFGAGVYVWGDGAGRNKPGVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 3
AD116817
ID AD116817 standard; protein; 855 AA.
AC AD116817;
XX
XX DT 15-APR-2004 (first entry)
XX
XX DE Human NOXV protein homologue SegID 353.
XX
XX KW human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;
XX
XX KW inflammation; autoimmune disorder; allergy; blood disorder;
XX
XX KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
XX
XX KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
XX
XX KW Alzheimer's disease; infection; str.
XX
XX OS Homo sapiens.
XX
XX PN WO200268649-A2.
XX
XX PD 06-SEP-2002.
XX

```

PF 31-JAN-2002; 2002W0-US002785.  
XX  
XX 31-JAN-2001; 2001US-0265395P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266767P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 07-FEB-2001; 2001US-0267057P.  
PR 08-FEB-2001; 2001US-0267459P.  
PR 09-FEB-2001; 2001US-0267823P.  
PR 15-FEB-2001; 2001US-0268974P.  
PR 26-FEB-2001; 2001US-0271664P.  
PR 27-FEB-2001; 2001US-0271839P.  
PR 27-FEB-2001; 2001US-0271855P.  
PR 02-MAR-2001; 2001US-0272788P.  
PR 02-MAR-2001; 2001US-0273046P.  
PR 14-MAR-2001; 2001US-0275925P.  
PR 14-MAR-2001; 2001US-0275947P.  
PR 14-MAR-2001; 2001US-0275950P.  
PR 14-MAR-2001; 2001US-0275989P.  
PR 15-MAR-2001; 2001US-0276448P.  
PR 15-MAR-2001; 2001US-0276450P.  
PR 16-MAR-2001; 2001US-0276397P.  
PR 16-MAR-2001; 2001US-0276768P.  
PR 20-MAR-2001; 2001US-0278652P.  
PR 26-MAR-2001; 2001US-0278778P.  
PR 26-MAR-2001; 2001US-0278778P.  
PR 29-MAR-2001; 2001US-0279882P.  
PR 29-MAR-2001; 2001US-0279884P.  
PR 30-MAR-2001; 2001US-0280147P.  
PR 11-APR-2001; 2001US-0282992P.  
PR 11-APR-2001; 2001US-0283083P.  
PR 20-APR-2001; 2001US-0285133P.  
PR 23-APR-2001; 2001US-0285749P.  
PR 03-MAY-2001; 2001US-0286327P.  
PR 03-MAY-2001; 2001US-0288504P.  
PR 29-MAY-2001; 2001US-0294047P.  
PR 30-MAY-2001; 2001US-0294473P.  
PR 08-JUN-2001; 2001US-0296964P.  
PR 18-JUN-2001; 2001US-0298959P.  
PR 18-JUN-2001; 2001US-0299334P.  
PR 11-AUG-2001; 2001US-0312020P.  
PR 16-AUG-2001; 2001US-0312889P.  
PR 16-AUG-2001; 2001US-0312908P.  
PR 21-AUG-2001; 2001US-0313390P.  
PR 28-AUG-2001; 2001US-0315470P.  
PR 31-AUG-2001; 2001US-0316447P.  
PR 07-SEP-2001; 2001US-0318115P.  
PR 07-SEP-2001; 2001US-0318118P.  
PR 12-SEP-2001; 2001US-0318740P.  
PR 19-SEP-2001; 2001US-0323379P.  
PR 18-OCT-2001; 2001US-0330245P.  
PR 18-OCT-2001; 2001US-0330308P.  
PR 14-NOV-2001; 2001US-0332701P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Therneer VT, Spytek KA, Zernhuen BD, Patturajan M, Shinkets RA,  
PI Li L, Gangolli EA, Padigaru M, Anderson DM, Rastelli L, Miller CE;  
PI Furtak VL, Tappier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CZA,  
PI Furtak K, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK, Burgess CE;  
XX  
XX WPI; 2002-706998/76.  
XX  
XX  
XX New NOVX polypeptides and nucleic acids, useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX  
XX  
XX Disclosure; SEQ ID NO 353; 1498bp; English.

CC This invention relates to a novel nucleic acids, and encoded polypeptides  
CC thereof, which have properties related to the stimulation of biochemical  
CC or physiological responses in a cell, tissue, organ or organism.  
CC Specifically, it refers to the use of biologically active fragments for  
CC diagnostic and prognostic assays and furthermore in the treatment of  
CC diverse pathological conditions. The present invention describes novel  
CC human and murine NOVX proteins, as well as methods to modulate their  
CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
CC The NOVX polypeptides, polynucleotides and antibodies are useful in  
CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,  
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
CC treating or preventing diseases such as inflammation, autoimmune  
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
CC (AIDS), obesity, asthma, immunoglobulin (IgA) nephropathy, cistrosis,  
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
CC and epilepsy. Accordingly, these molecules have many activities including  
CC cytostatic, cardiant, antiinflammatory, immunosuppressive, antiallergic,  
CC hemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,  
CC antiaslomatic, nephrotropic, antiarthritic, hepatotropic,  
CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,  
CC relaxant and anticonvulsant. In addition, they are useful in screening  
CC assays to identify small molecules that modulate or inhibit, for example,  
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
CC used as in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein  
CC of the invention.  
XX  
XX  
XX Sequence 855 AA;  
SQ  
Query Match 99.5%; Score 4659; DB 5; Length 855;  
Best Local Similarity 99.0%; Pred. No. 9.4e-308;  
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MGSDRRKGGGKPKDAGLKYNSRHEKVGLEEGVEFLPVNNVKKYKGGPWWVLA 60  
DB 1 MGSDRRKGGGKPKDAGLKYNSRHEKVGLEEGVEFLPVNNVKKYKGGPWWVLA 60  
QY VILGLILVILGIGFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENNSTEVSLSKY 120  
DB VILGLILVILGIGFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENNSTEVSLSKY 120  
QY 121 KDLKILYSGVPLGPGYKHESAVTAFSEGSVIAVYSEFSIPQHLVEBARVVAERVVM 180  
DB 121 KDLKILYSGVPLGPGYKHESAVTAFSEGSVIAVYSEFSIPQHLVEBARVVAERVVM 180  
QY 121 KDLKILYSGVPLGPGYKHESAVTAFSEGSVIAVYSEFSIPQHLVEBARVVAERVVM 180  
DB 121 KDLKILYSGVPLGPGYKHESAVTAFSEGSVIAVYSEFSIPQHLVEBARVVAERVVM 180  
QY 181 LPPRASLSKSFVTSVVAEPDTSKTVQRTQDNCSEGLHARGVELMFTTGPDPSPYPA 240  
DB 181 LPPRASLSKSFVTSVVAEPDTSKTVQRTQDNCSEGLHARGVELMFTTGPDPSPYPA 240  
QY 241 HARCQALRGDADSVLTFRSPDLASCDERGSDLVTYNTLSPMERHALVOLCGTYPYS 300  
DB 241 HARCQALRGDADSVLTFRSPDLASCDERGSDLVTYNTLSPMERHALVOLCGTYPYS 300  
QY 241 HARCQALRGDADSVLTFRSPDLASCDERGSDLVTYNTLSPMERHALVOLCGTYPYS 300  
DB 241 HARCQALRGDADSVLTFRSPDLASCDERGSDLVTYNTLSPMERHALVOLCGTYPYS 300  
QY 301 YNLTFFSSQNVLLITLTNTERRHGFEPATFPQLPMSSCGGLRAQGFNSPYPGHY 360  
DB 301 YNLTFFSSQNVLLITLTNTERRHGFEPATFPQLPMSSCGGLRAQGFNSPYPGHY 360  
QY 361 PNVIDCTWNIIEVNNQVKKVRFKFLLEBPGVAGTCPKDYVINEKCYGERSQCVYVS 420  
DB 361 PNVIDCTWNIIEVNNQVKKVRFKFLLEBPGVAGTCPKDYVINEKCYGERSQCVYVS 420  
QY 421 NSNKITVRPHSDSYTDGFLAFLYISYSSDPCEGQFTCKTGRCIRKELRDGMADCTTH 480  
DB 421 NSNKITVRPHSDSYTDGFLAFLYISYSSDPCEGQFTCKTGRCIRKELRDGMADCTTH 480  
QY 481 SDELNCSCAQHQFTCKNFKCPLFWVCDVSDGNSDBQGSCEPAQTFRSGNGLSK 540  
DB 481 SDELNCSCAQHQFTCKNFKCPLFWVCDVSDGNSDBQGSCEPAQTFRSGNGLSK 540  
QY 541 SQQCNKDKDCGSDSDASCPKVVVYCTGHTYRCNLGLCLSKGNPECDGKEDSDSDSK 600  
DB 541 SQQCNKDKDCGSDSDASCPKVVVYCTGHTYRCNLGLCLSKGNPECDGKEDSDSDSK 600

QY 601 DCDGSLRFTFOARVVGSTDADEGEMPMQVSLHALGSHICGASLISPNMLVSAHACYID 660  
Db 601 DCDGSLRFTFOARVVGSTDADEGEMPMQVSLHALGSHICGASLISPNMLVSAHACYID 660  
QY 661 DRGRVSDPTQMTAFGLIHDSQSRAPQVQRRLKRIISHPPDFTFDYDIALLELEKP 720  
Db 661 DRGRVSDPTQMTAFGLIHDSQSRAPQVQRRLKRIISHPPDFTFDYDIALLELEKP 720  
QY 721 AEVSSMRPITCLPASHVFPFGKAIWYTGWGHYGGGALLIQKGRIRVNOTTCENLL 780  
Db 721 AEVSSMRPITCLPASHVFPFGKAIWYTGWGHYGGGALLIQKGRIRVNOTTCENLL 780  
QY 781 PQDITPRMTCVGFSGVDSQGDSPGFLSSVEADGRIFGAGVYMWGDCAGRNKPGVYT 840  
Db 781 PQDITPRMTCVGFSGVDSQGDSPGFLSSVEADGRIFGAGVYMWGDCAGRNKPGVYT 840  
QY 841 RLPLFRDMIKENTGV 855  
Db 841 RLPLFRDMIKENTGV 855  
RESULT 4  
AD116883  
ID AD116883 standard; protein; 855 AA.  
XX  
AC AD116883;  
XX  
DT 15-APR-2004 (first entry)  
XX  
DE Human NOVX protein homologue seqid 419.  
XX  
KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;  
KW inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig) A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; str.  
XX  
OS Homo sapiens.  
XX  
PN W0200268649-A2.  
XX  
PD 06-SEP-2002.  
XX  
PF 31-JAN-2002; 2002W0-US002785.  
XX  
PR 31-JAN-2001; 2001US-0265395P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266767P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 07-FEB-2001; 2001US-0267057P.  
PR 08-FEB-2001; 2001US-0267459P.  
PR 09-FEB-2001; 2001US-0267823P.  
PR 15-FEB-2001; 2001US-0268974P.  
PR 26-FEB-2001; 2001US-0271664P.  
PR 27-FEB-2001; 2001US-0271839P.  
PR 02-MAR-2001; 2001US-0272788P.  
PR 02-MAR-2001; 2001US-0273046P.  
PR 14-MAR-2001; 2001US-0275925P.  
PR 14-MAR-2001; 2001US-0275947P.  
PR 14-MAR-2001; 2001US-0275950P.  
PR 14-MAR-2001; 2001US-0275989P.  
PR 15-MAR-2001; 2001US-0276448P.  
PR 15-MAR-2001; 2001US-0276450P.  
PR 16-MAR-2001; 2001US-0276397P.  
PR 16-MAR-2001; 2001US-0276768P.  
PR 20-MAR-2001; 2001US-0278652P.  
PR 26-MAR-2001; 2001US-0278775P.  
PR 26-MAR-2001; 2001US-0278778P.  
PR 29-MAR-2001; 2001US-0279882P.

PR 29-MAR-2001; 2001US-0279884P.  
PR 30-MAR-2001; 2001US-0280147P.  
PR 11-APR-2001; 2001US-0282992P.  
PR 11-APR-2001; 2001US-0283083P.  
PR 20-APR-2001; 2001US-0285133P.  
PR 20-APR-2001; 2001US-0285749P.  
PR 03-MAY-2001; 2001US-0288337P.  
PR 03-MAY-2001; 2001US-0288504P.  
PR 29-MAY-2001; 2001US-0294047P.  
PR 30-MAY-2001; 2001US-0294473P.  
PR 08-JUN-2001; 2001US-0296964P.  
PR 18-JUN-2001; 2001US-0298959P.  
PR 19-JUN-2001; 2001US-0299324P.  
PR 13-AUG-2001; 2001US-0312020P.  
PR 16-AUG-2001; 2001US-0312889P.  
PR 16-AUG-2001; 2001US-0312908P.  
PR 21-AUG-2001; 2001US-0313350P.  
PR 28-AUG-2001; 2001US-0315470P.  
PR 31-AUG-2001; 2001US-0316447P.  
PR 07-SEP-2001; 2001US-0318115P.  
PR 07-SEP-2001; 2001US-0318118P.  
PR 12-SEP-2001; 2001US-0318740P.  
PR 19-SEP-2001; 2001US-0323379P.  
PR 18-OCT-2001; 2001US-0330245P.  
PR 18-OCT-2001; 2001US-0330308P.  
PR 14-NOV-2001; 2001US-0332701P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Tchernev VT, Spytek RA, Zehrhusen BD, Paturajan M, Shinkets RA,  
XX Li L, Gangoli EA, Padigar M, Anderson DW, Rascelli L, Miller CE,  
XX Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wascell AR, Pena CRA,  
XX Furtak K, Grosse WM, Alsebrook JP, Lepley DM, Rieger DK, Burgess CE,  
XX WPI; 2002-706998/76.  
XX  
XX New NOVX polypeptides and nucleic acids, useful for preventing or  
XX treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
XX atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
XX pharmacogenomics.  
XX  
XX Disclosure; SEQ ID NO 419; 1498bp; English.  
XX  
XX This invention relates to a novel nucleic acids, and encoded polypeptides  
XX thereof, which have properties related to the stimulation of biochemical  
XX or physiological responses in a cell, tissue, organ or organism.  
XX Specifically, it refers to the use of biologically active fragments for  
XX diagnostic and prognostic assays and furthermore in the treatment of  
XX diverse pathological conditions. The present invention describes novel  
XX human and murine NOVX proteins, as well as methods to modulate their  
XX expression using antisense oligos, ribozymes and peptide nucleic acids.  
XX The NOVX polypeptides, polynucleotides and antibodies are useful in  
XX treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,  
XX atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
XX treating or preventing diseases such as inflammation, autoimmune  
XX disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
XX (AIDS), obesity, asthma, immunoglobulin (Ig) A nephropathy, cirrhosis,  
XX arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
XX and epilepsy. Accordingly, these molecules have many activities including  
XX cytoskeletal, cardiac, antiinflammatory, immunosuppressive, antiallergic,  
XX haemostatic, anti-HIV, antidiabetic, antiatherosclerotic, anorectic,  
XX antiasthmatic, nephrotropic, antiarthritic, hepatotropic,  
XX neuroprotective, neurotropic, antibacterial, vitruide, antiparasitic,  
XX relaxant and anticonvulsant. In addition, they are useful in screening  
XX assays to identify small molecules that modulate or inhibit, for example,  
XX neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
XX used as in chromosome mapping, tissue typing, preventive medicine and  
XX pharmacogenomics. This polypeptide is a homologue of a human NOVX protein  
XX of the invention.  
XX  
XX Sequence 855 AA;  
SQ

Query Match

99.5%; Score 4659; DB 5; Length 855;



Best Local Similarity 99.6%; Pred. No. 9.4e-308;  
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MGSDRAKGGGGPPDFGAGLKYNRRHKVNGLEGEVFLPNNVKKYKEKPGGVVLA 60
Db 1 MGSDRAKGGGGPPDFGAGLKYNRRHKVNGLEGEVFLPNNVKKYKEKPGGVVLA 60
QY 61 VLIGLLVLLIGLIGFLVWHLQYRDVVOVKVKNYKRIINENFVDAEENSNTSEFVSLASKV 120
Db 61 VLIGLLVLLIGLIGFLVWHLQYRDVVOKVFGYKRIINENFVDAEENSNTSEFVSLASKV 120
QY 121 KDAKLKLYSGVPPFLGPHKESAVTAFSEGSYIATYNSSEFSLPQHLVEAEKVAEEVVM 180
Db 121 KDAKLKLYSGVPPFLGPHKESAVTAFSEGSYIATYNSSEFSLPQHLVEAEKVAEEVVM 180
QY 181 LPPRARSLSKSVVTSVVAFPDTSKTVORTDNSSCSFGLHARGVLEKMPFTTGPFPDSDPYPA 240
Db 181 LPPRARSLSKSVVTSVVAFPDTSKTVORTDNSSCSFGLHARGVLEKMPFTTGPFPDSDPYPA 240
QY 241 HARCQMLRGDADSVLSLTFRSPDLASCDEKSGDLVTYNTLSPEMETHALVOLCGTYPPS 300
Db 241 HARCQMLRGDADSVLSLTFRSPDLASCDEKSGDLVTYNTLSPEMETHALVOLCGTYPPS 300
QY 301 YNLTFFSSQNVLLITLTNTERRHPGEBATFFQLPRNMSCCGRLRKAQGTFRNSYRGGHY 360
Db 301 YNLTFFSSQNVLLITLTNTERRHPGEBATFFQLPRNMSCCGRLRKAQGTFRNSYRGGHY 360
QY 361 PPNIDCTWNIIEVPNNQVKVRFKFFYLLEPVPAGTCKPVVEINGEKYGCEGSEQFVYTS 420
Db 361 PPNIDCTWNIIEVPNNQVKVRFKFFYLLEPVPAGTCKPVVEINGEKYGCEGSEQFVYTS 420
QY 421 NSNKITVRFHSDQSYTDGTGLAEYLSYSDSPCGQFCTCRGTCIRKELRCGWDCTDH 480
Db 421 NSNKITVRFHSDQSYTDGTGLAEYLSYSDSPCGQFCTCRGTCIRKELRCGWDCTDH 480
QY 481 SDELNCSGDAGHOTCKNKKFKPLFWICDSYNDGSDNSDEQSCSPQOTRCSNGKCLSK 540
Db 481 SDELNCSGDAGHOTCKNKKFKPLFWICDSYNDGSDNSDEQSCSPQOTRCSNGKCLSK 540
QY 541 SQQNGKDDCGDSDSEASCPKVNVTCTKHYRCLANGCLSKNGPECDGKEDCGSDGDEK 600
Db 541 SQQNGKDDCGDSDSEASCPKVNVTCTKHYRCLANGCLSKNGPECDGKEDCGSDGDEK 600
QY 601 DCDGGLSFTROARVVGTDADGEGMPQVSLHALGGHICGASLLSPNMLVSAHACYID 660
Db 601 DCDGGLSFTROARVVGTDADGEGMPQVSLHALGGHICGASLLSPNMLVSAHACYID 660
QY 661 DRGRYSDPTQWTFPLGLHDOSASAPGVORRLKRIISHPFNDFTFYDIALLELEKP 720
Db 661 DRGRYSDPTQWTFPLGLHDOSASAPGVORRLKRIISHPFNDFTFYDIALLELEKP 720
QY 721 AEYSMWRTICLPDASHVFPAGKAIWYTGWHTQYGGTGALILQKGEIRVINOTTGENTL 780
Db 721 AEYSMWRTICLPDASHVFPAGKAIWYTGWHTQYGGTGALILQKGEIRVINOTTGENTL 780
QY 781 PQQITPRMVCVGLSGVDSCQDGSGLSSVEADGRIFGAGVVSWDGCAGRNKPQVYT 840
Db 781 PQQITPRMVCVGLSGVDSCQDGSGLSSVEADGRIFGAGVVSWDGCAGRNKPQVYT 840
QY 841 RLPLFRMIKENTGV 855
Db 841 RLPLFRMIKENTGV 855
```

## RESULT 5

AD116876  
ID AD116876 standard; protein; 855 AA.

XX AC AD116876;

XX DT 15-APR-2004 (first entry)

XX DE Human NOXV protein homologue SegID 412.

XX human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;  
KW inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; str.

XX Homo sapiens.

PN W0200268649-A2.

XX 06-SEP-2002.

XX 31-JAN-2002; 2002W0-US002785.

XX 31-JAN-2001; 2001US-0265395P.

XX 31-JAN-2001; 2001US-0265412P.

XX 31-JAN-2001; 2001US-0265514P.

XX 02-FEB-2001; 2001US-0265517P.

XX 05-FEB-2001; 2001US-0266406P.

XX 07-FEB-2001; 2001US-0266767P.

XX 07-FEB-2001; 2001US-0266975P.

XX 08-FEB-2001; 2001US-0267057P.

XX 09-FEB-2001; 2001US-0267459P.

XX 15-FEB-2001; 2001US-0267823P.

XX 26-FEB-2001; 2001US-0271664P.

XX 27-FEB-2001; 2001US-0271855P.

XX 02-MAR-2001; 2001US-0271939P.

XX 02-MAR-2001; 2001US-0272788P.

XX 14-MAR-2001; 2001US-0273046P.

XX 14-MAR-2001; 2001US-0275325P.

XX 14-MAR-2001; 2001US-0275947P.

XX 14-MAR-2001; 2001US-0275950P.

XX 15-MAR-2001; 2001US-0275989P.

XX 15-MAR-2001; 2001US-0276448P.

XX 15-MAR-2001; 2001US-0276450P.

XX 16-MAR-2001; 2001US-0276397P.

XX 20-MAR-2001; 2001US-027678P.

XX 20-MAR-2001; 2001US-0278652P.

XX 26-MAR-2001; 2001US-0278775P.

XX 26-MAR-2001; 2001US-0278778P.

XX 29-MAR-2001; 2001US-0279682P.

XX 29-MAR-2001; 2001US-0279684P.

XX 30-MAR-2001; 2001US-0280147P.

XX 11-APR-2001; 2001US-0282992P.

XX 11-APR-2001; 2001US-0283083P.

XX 20-APR-2001; 2001US-0285133P.

XX 23-APR-2001; 2001US-0285749P.

XX 03-MAY-2001; 2001US-0288327P.

XX 03-MAY-2001; 2001US-0288504P.

XX 29-MAY-2001; 2001US-0294047P.

XX 30-MAY-2001; 2001US-0294473P.

XX 08-JUN-2001; 2001US-0296364P.

XX 18-JUN-2001; 2001US-0298959P.

XX 19-JUN-2001; 2001US-0299324P.

XX 13-AUG-2001; 2001US-0312020P.

XX 16-AUG-2001; 2001US-0312889P.

XX 16-AUG-2001; 2001US-0312908P.

XX 21-AUG-2001; 2001US-0313390P.

XX 28-AUG-2001; 2001US-0315470P.

XX 31-AUG-2001; 2001US-0316447P.

XX 07-SEP-2001; 2001US-0318115P.

XX 07-SEP-2001; 2001US-0318118P.

XX 12-SEP-2001; 2001US-0318740P.

XX 19-SEP-2001; 2001US-0323379P.

XX 18-OCT-2001; 2001US-0330245P.

XX 18-OCT-2001; 2001US-0330308P.

XX 14-NOV-2001; 2001US-0332701P.

(CURA-) CURAGEN CORP.

XX Tchernev VT, Spytek KA, Zernusen BD, Paturajan M, Shinkets RA;

PI Li L, Gangoli EA, Padigaru M, Anderson DM, Raselli L, Miller CE;  
PI Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CRA;  
PI Purtek K, Grosse WM, Alsebrook JP, Lepley DM, Rieger DK, Burgess CE;  
XX MPI; 2002-706998/76.  
XX  
XX New NOVX polypeptides and nucleic acids, useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.

XX  
XX Disclosure; SEQ ID NO 412; 1498bp; English.

XX  
XX This invention relates to a novel nucleic acid, and encoded polypeptides  
XX thereof, which have properties related to the stimulation of biochemical  
XX or physiological responses in a cell, tissue, organ or organism.  
XX Specifically, it refers to the use of biologically active fragments for  
XX diagnostic and prognostic assays and furthermore in the treatment of  
XX diverse pathological conditions. The present invention describes novel  
XX human and murine NOVX proteins, as well as methods to modulate their  
XX expression using antisense oligos, ribozymes and peptide nucleic acids.  
XX The NOVX polypeptides, polynucleotides and antibodies are useful in  
XX treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,  
XX atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
XX treating or preventing diseases such as inflammation, autoimmune  
XX disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
XX (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,  
XX arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
XX and epilepsy. Accordingly, these molecules have many activities including  
XX cytoskeletal, cardiac, anti-inflammatory, immunosuppressive, antiallergic,  
XX haemostatic, anti-HIV, antidiabetic, antierosive, anorectic,  
XX antiasthmatic, nephroprotective, antirheumatic, hepatoprotective,  
XX neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,  
XX relaxant and anticonvulsant. In addition, they are useful in screening  
XX assays to identify small molecules that modulate or inhibit, for example,  
XX neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
XX used as in chromosome mapping, tissue typing, preventive medicine and  
XX pharmacogenomics. This polypeptide is a homologue of a human NOVX protein  
XX of the invention.

XX  
XX Sequence 855 AA;

Query Match 99.5%; Score 4659; DB 5; Length 855;

Best Local Similarity 99.6%; Pred. No. 9.4e-308;

Matches 855; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSDBARKGGGGPPDFAGLKYNSRHRKVNGLBEVFLPVNNVKKYKGGKGRWVLLAA 60  
Db 1 MGSDBARKGGGGPPDFAGLKYNSRHRKVNGLBEVFLPVNNVKKYKGGKGRWVLLAA 60  
QY 61 VLIGLLVLLIGLIGLVNHLQYRDVQVKNKGNKRNINENPVDAKENSNSTEFSLASKV 120  
Db 61 VLIGLLVLLIGLIGLVNHLQYRDVQVKNKGNKRNINENPVDAKENSNSTEFSLASKV 120  
QY 121 KDALKLKLYSGVFLGPRHKSAYTAFSEGSVIAIYVMESEFIPOLVBEAEKVMAEKRVVM 180  
Db 121 KDALKLKLYSGVFLGPRHKSAYTAFSEGSVIAIYVMESEFIPOLVBEAEKVMAEKRVVM 180  
QY 181 LPPRARSLSKSFVVTSVVAFPTDSKTQVORTQDNCSFGILHARGVELMFTTGPDPSPYPA 240  
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTQVORTQDNCSFGILHARGVELMFTTGPDPSPYPA 240  
QY 241 HARGQVALLRGDADSVLSLTPRSFPLASCDERGSPLVYVNTLSMEPHALVOLCGTTPPS 300  
Db 241 HARGQVALLRGDADSVLSLTPRSFPLASCDERGSPLVYVNTLSMEPHALVOLCGTTPPS 300  
QY 301 YNLFHSSQNVLLTTLTNTERRHGPFEATPFQLPKMSCGGRLLKKAQGFNSPYPGHY 360  
Db 301 YNLFHSSQNVLLTTLTNTERRHGPFEATPFQLPKMSCGGRLLKKAQGFNSPYPGHY 360  
QY 361 PPNDCTWNIIEVPNNQHVKKAFKFFYLLBPGVAGTCPKDVEINGEKCYCGERSQFVVT 420  
Db 361 PPNDCTWNIIEVPNNQHVKKAFKFFYLLBPGVAGTCPKDVEINGEKCYCGERSQFVVT 420

QY 421 NSNKITVRFHSDSDYTDTGFLAEYLSYDSSDPGPGQFTCTGRCIRKELRCDDWADCTDH 480  
Db 421 NSNKITVRFHSDSDYTDTGFLAEYLSYDSSDPGPGQFTCTGRCIRKELRCDDWADCTDH 480  
QY 481 SDELNCSGADAGHPTCKNKKCPPLFWVCDSVNDGSDNSDQSCSPAQTRCSNGKLSK 540  
Db 481 SDELNCSGADAGHPTCKNKKCPPLFWVCDSVNDGSDNSDQSCSPAQTRCSNGKLSK 540  
QY 541 SQQCNGDDDCGSDASCKRNVVYCTKATYRCLNGLCLSKNPECDGDCSDGSD 600  
Db 541 SQQCNGDDDCGSDASCKRNVVYCTKATYRCLNGLCLSKNPECDGDCSDGSD 600  
QY 601 DCDGSLSPTRQARVVGATDAGEMPPOVSLHALGQGHICGASLISPMVLVSAHGYID 660  
Db 601 DCDGSLSPTRQARVVGATDAGEMPPOVSLHALGQGHICGASLISPMVLVSAHGYID 660  
QY 661 DRGRYSPTQWTAFLGLHDQSORSAFGVERBLKRIISHPFNDFTDYDIALLEK 720  
Db 661 DRGRYSPTQWTAFLGLHDQSORSAFGVERBLKRIISHPFNDFTDYDIALLEK 720  
QY 721 AYSVMWRPCLDPAHVFPAGKAIWYTGWHGYOYGTGALLIOKGRIRINOTTCENL 780  
Db 721 AYSVMWRPCLDPAHVFPAGKAIWYTGWHGYOYGTGALLIOKGRIRINOTTCENL 780  
QY 781 POOTTPRMCGVFLSGGVSDSCGDSGGLSVSEADGRIFGAVVWMDGACGRKPGVYT 840  
Db 781 POOTTPRMCGVFLSGGVSDSCGDSGGLSVSEADGRIFGAVVWMDGACGRKPGVYT 840  
QY 841 RLPLFRDMIKENTGV 855  
Db 841 RLPLFRDMIKENTGV 855

RESULT 6

ID ADN39867 standard; protein; 855 AA.

ADN39867;

17-JUN-2004 (first entry)

Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO: C237.

Human, differential expression; cancer; angiogenic disorder;  
fibrotic disorder; psoriasis; leukaemia; heart disease; atherosclerosis;  
inflammatory disease; autoimmune disease;  
retinal neovascularisation syndrome; scarring; uterine fibroid;  
detection; diagnosis; prognosis; drug screening; drug targeting;  
wound healing; contraception; cytostatic; cardiac; immunomodulatory;  
vulnerability; gene therapy; vaccine.

Homo sapiens.

MO2003042661-A2.

22-MAY-2003.

13-NOV-2002; 2002MO-US036810.

13-NOV-2001; 2001US-0350666P.

21-NOV-2001; 2001US-0332464P.

29-NOV-2001; 2001US-0334383P.

03-DEC-2001; 2001US-0335394P.

14-DEC-2001; 2001US-0340376P.

08-JAN-2002; 2002US-0347211P.

10-JAN-2002; 2002US-0347349P.

08-FEB-2002; 2002US-0355250P.

13-FEB-2002; 2002US-0356714P.

20-FEB-2002; 2002US-0359077P.

29-MAR-2002; 2002US-0368099P.

04-APR-2002; 2002US-0370110P.

12-APR-2002; 2002US-0372246P.

PR 05-JUN-2002; 2002US-0386614P.  
PR 16-JUL-2002; 2002US-0396839P.  
PR 22-JUL-2002; 2002US-0397775P.  
PR 23-JUL-2002; 2002US-0397845P.  
PR 09-SEP-2002; 2002US-0409450P.  
XX  
PA (EOSB-) EOS BIOTECHNOLOGY INC.  
PI Afar D, Aziz N, Ginsburg WM, Gish KC, Glyme R, Hevezi PA;  
PI Mack DH, Murray R, Watson SR, Wilson KE, Zlocnik A;  
XX  
XX WPI; 2003-468649/44.  
DR N-PSDB; ADN39650.  
XX  
PT Determining the presence or absence of a pathological cell in a patient,  
PT useful for diagnosing, prognosing or treating cancer, comprises detecting  
PT a nucleic acid in a biological sample.  
XX  
XX  
PS Claim 12; SEQ ID NO C237; 1385bp; English.  
XX  
XX The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
CC whose expression is upregulated or downregulated in specific cancers or  
CC other diseases such as angiogenic or fibrotic disorders, and to methods  
CC of determining the presence or absence of a pathological cell in a  
CC patient by detecting a nucleic acid at least 80% identical to those of  
CC the invention or by detecting a polypeptide of the invention. The  
CC invention also relates to expression vectors and host cells comprising a  
CC nucleic acid of the invention; antibodies which specifically bind a  
CC polypeptide of the invention; use of such antibodies for drug targeting;  
CC and methods of screening for modulators of activity or expression of the  
CC polypeptides and nucleic acids. The nucleic acids, polypeptides,  
CC antibodies and methods are useful for diagnosing, prognosing and treating  
CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
CC neovascularisation syndromes, scarring and ureteric fibrosis. They may  
CC also be useful in wound healing and in contraception. The present  
CC sequence represents a polypeptide of the invention.  
XX  
XX Sequence 855 AA;  
SQ  
Query Match 99.5%; Score 4659; DB 7; Length 855;  
Best Local Similarity 99.6%; Pred. No. 9.4e-308;  
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MGSDBARKGGGKDPFAGLKYNSRHEKVNGLBEGVEFLPVNNVKYKKGPGRWVTLAA 60  
DB 1 MGSDBARKGGGKDPFAGLKYNSRHEKVNGLBEGVEFLPVNNVKYKKGPGRWVTLAA 60  
QY 61 VILGLLVLLIGIGLVVHLLQYRDYRVOVKXGKRYRINENFVDAYENSSTEFYSLASKV 120  
DB 61 VILGLLVLLIGIGLVVHLLQYRDYRVOVKXGKRYRINENFVDAYENSSTEFYSLASKV 120  
QY 121 KDALKLLYSVGPFLGPRYHKSAAVTAFSEGSVIATYMESEFSIPQHLVEARVMAEERVVM 180  
DB 121 KDALKLLYSVGPFLGPRYHKSAAVTAFSEGSVIATYMESEFSIPQHLVEARVMAEERVVM 180  
QY 181 LPPRARSLSKSFVYTSVVAFPDTSKTYQRTODNSCSFGLHARGVELMFTTGPEDSPYPA 240  
DB 181 LPPRARSLSKSFVYTSVVAFPDTSKTYQRTODNSCSFGLHARGVELMFTTGPEDSPYPA 240  
QY 241 HARCQMLRQDADSVLSLTRSPFLASCDEGSGSLVTVNTTSLSPMEHALVOLCGTTPPS 300  
DB 241 HARCQMLRQDADSVLSLTRSPFLASCDEGSGSLVTVNTTSLSPMEHALVOLCGTTPPS 300  
QY 301 YNLTFFSSQVLLITLTTNTERRRPGFEATFPOLPRMSSCGGRIRKKAQGFNSPYPGHY 360  
DB 301 YNLTFFSSQVLLITLTTNTERRRPGFEATFPOLPRMSSCGGRIRKKAQGFNSPYPGHY 360  
QY 361 PPNIIDCTWNIENVNQHVKVRFKFFYLLEBGPAGTCPKDYVEINSEKYGESRQFVYTS 420  
DB 361 PPNIIDCTWNIENVNQHVKVRFKFFYLLEBGPAGTCPKDYVEINSEKYGESRQFVYTS 420  
QY 421 NSNKITVRFHSDOSYTDYTGFLAETLSYDSDPCPGQFTCRTRCIRKELRCDCMACDCTDH 480

DB 421 NSNKITVRFHSDOSYTDYTGFLAETLSYDSDPCPGQFTCRTRCIRKELRCDCMACDCTDH 480  
QY 481 SDELNCSCAAGHOFCTCKNFKCKPLFWVCDSDVNDGNSDQGSCEPAQTFRCNKGCLSK 540  
DB 481 SDELNCSCAAGHOFCTCKNFKCKPLFWVCDSDVNDGNSDQGSCEPAQTFRCNKGCLSK 540  
QY 541 SQQCNKGDDCGSDSDASCEKXNVVYCTKATYRCNLGLCLSKNPECDGKEDCSDSDEK 600  
DB 541 SQQCNKGDDCGSDSDASCEKXNVVYCTKATYRCNLGLCLSKNPECDGKEDCSDSDEK 600  
QY 601 DCCCGLRSTFRQARVYVGTGTDADGEMPMQVSLHALQGHICGASLISPMVLVSAACCTID 660  
DB 601 DCCCGLRSTFRQARVYVGTGTDADGEMPMQVSLHALQGHICGASLISPMVLVSAACCTID 660  
QY 661 DRGFRYSDFQWTAFLGHDQSORSAFGVERLKRILISHPFNDFTFDIALLELEXP 720  
DB 661 DRGFRYSDFQWTAFLGHDQSORSAFGVERLKRILISHPFNDFTFDIALLELEXP 720  
QY 721 ABEYSSWVRPILCLPDASHVPAGRAIWTGNGHTQYGGTGAIIQKGEIRVINGTCENLL 780  
DB 721 ABEYSSWVRPILCLPDASHVPAGRAIWTGNGHTQYGGTGAIIQKGEIRVINGTCENLL 780  
QY 781 PQQITRRMVCVGLSGVDSCQDSSGGLSVADGRIRGAGVYVSGDCAKNNKGVYT 840  
DB 781 PQQITRRMVCVGLSGVDSCQDSSGGLSVADGRIRGAGVYVSGDCAKNNKGVYT 840  
QY 841 RLPLFRDWIKENTGV 855  
DB 841 RLPLFRDWIKENTGV 855  
RESULT 7  
ADN04754  
ID ADN04754 standard; protein; 855 AA.  
XX  
AC ADN04754;  
XX  
DT 01-JUL-2004 (first entry)  
XX  
DE Antipsoriatic protein sequence #558.  
XX  
XX  
KW antipsoriatic; gene therapy; psoriasis; diagnosis.  
XX  
OS Homo sapiens.  
XX  
PN WC02004028479-A2.  
XX  
PD 08-APR-2004.  
XX  
PF 25-SEP-2003; 2003MO-US030907.  
XX  
PR 25-SEP-2002; 2002US-0414006P.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Boday S, Clark H, Jackman J, Schoenfeld J, Williams PW, Wood WI;  
PI Wu TD;  
XX  
XX WPI; 2004-305105/28.  
DR N-PSDB; ADN04753.  
XX  
PT New PRO nucleic acid or polypeptide, useful for preparing a  
PT pharmaceutical composition for diagnosing or treating psoriasis in a  
PT mammal.  
XX  
XX Claim 9; SEQ ID NO 1148; 3069bp; English.  
XX  
XX The invention relates to novel polynucleotide and polypeptides for  
XX treating psoriasis or a sequence having at least 80% identity to the  
XX above sequences. The nucleic acid is useful for preparing a composition  
XX for diagnosing or treating psoriasis in a mammal. This sequence  
XX corresponds to one of the polypeptides of the invention.

XX Sequence 855 AA:  
SQ  
Query Match 99.5%; Score 4659; DB 8; Length 855;  
Best Local Similarity 99.6%; Pred. No. 9,4e-308;  
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MGSDRAKGGGGPGDFGAGLKYNSRHEKRNGLIEGVEFLPVNNYKYEKQKPGWVYLA 60  
DB 1 MGSDRAKGGGGPGDFGAGLKYNSRHEKRNGLIEGVEFLPVNNYKYEKQKPGWVYLA 60  
QY 61 VLIGLLVLIGIFLVNHLQYRDVYQVKKNYRITNENFVDAEYNSNSTEYSLASKV 120  
DB 61 VLIGLLVLIGIFLVNHLQYRDVYQVKKNYRITNENFVDAEYNSNSTEYSLASKV 120  
QY 121 KDALKLYSGVYFPGFPHKSAVTAFFSGSYIATYWSFSLPQHLVEAEKVAEERYVM 180  
DB 121 KDALKLYSGVYFPGFPHKSAVTAFFSGSYIATYWSFSLPQHLVEAEKVAEERYVM 180  
QY 181 LPPRARSLSKSPVTVSVAFPTDSKTVOPTDQNSCSFGHARGVBLMFTTPGFPDSPYPA 240  
DB 181 LPPRARSLSKSPVTVSVAFPTDSKTVOPTDQNSCSFGHARGVBLMFTTPGFPDSPYPA 240  
QY 241 HARCQWALRGDADSVLSLTFPSFDLASCDESGSDLVTVYNTLSPEMHPALVOLGTYPPS 300  
DB 241 HARCQWALRGDADSVLSLTFPSFDLASCDESGSDLVTVYNTLSPEMHPALVOLGTYPPS 300  
QY 301 YNLTFFSSQNVLLTLITNTERRHHPGFAITFQLPKMSCCGRLRKAQGTNSPYRGHY 360  
DB 301 YNLTFFSSQNVLLTLITNTERRHHPGFAITFQLPKMSCCGRLRKAQGTNSPYRGHY 360  
QY 361 PPNIDCTMNTEVPNNQHVAFKFFYLEPVPAGTCKDVEJNGEKYCEKSGSFVYVS 420  
DB 361 PPNIDCTMNTEVPNNQHVAFKFFYLEPVPAGTCKDVEJNGEKYCEKSGSFVYVS 420  
QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSDPCEGFTCRGRCIRKELRCGWDADCTDH 480  
DB 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSDPCEGFTCRGRCIRKELRCGWDADCTDH 480  
QY 481 SDELNGSCDAGHQTCKNKKFCKPLFWCDSVNDGSDNSDEGSCCPAQTFRCNSGKCLSK 540  
DB 481 SDELNGSCDAGHQTCKNKKFCKPLFWCDSVNDGSDNSDEGSCCPAQTFRCNSGKCLSK 540  
QY 541 SQQNGDGDGSDGSDASCPKVNVTCTKHTYRCLNGCLSKNPECDGKEDCSGSDSEK 600  
DB 541 SQQNGDGDGSDGSDASCPKVNVTCTKHTYRCLNGCLSKNPECDGKEDCSGSDSEK 600  
QY 601 DCDGSLRFTQRAVVGSTDADEGEWPQVSLHALGQGHICGASLLSPNMLVSAHCYID 660  
DB 601 DCDGSLRFTQRAVVGSTDADEGEWPQVSLHALGQGHICGASLLSPNMLVSAHCYID 660  
QY 661 DRGRYSDDPTQWTAFLGLHDSQSAAPGVQRRRLKRIISHPFNDFTFYDIALLELEKP 720  
DB 661 DRGRYSDDPTQWTAFLGLHDSQSAAPGVQRRRLKRIISHPFNDFTFYDIALLELEKP 720  
QY 721 AEYSSMRPITCLPASHVFPAGKAIWYTGWHTYGGGALLQKGEIRVNTQTTCELL 780  
DB 721 AEYSSMRPITCLPASHVFPAGKAIWYTGWHTYGGGALLQKGEIRVNTQTTCELL 780  
QY 781 PQQITPMMVCYFSGVDSQQSDSGPLSVEADGRIFGAGVYSWGDCAGRKPKGYT 840  
DB 781 PQQITPMMVCYFSGVDSQQSDSGPLSVEADGRIFGAGVYSWGDCAGRKPKGYT 840  
QY 841 RLPLFRDWIKENTGV 855  
DB 841 RLPLFRDWIKENTGV 855

XX 15-Apr-2004 (first entry)  
DT Human NOX protein homologue SegID 420.  
XX  
DE human; NOX; cardiomyopathy; atherosclerosis; cancer; diabetes;  
XX inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
XX Alzheimer's disease; infection; str.  
OS Homo sapiens.  
XX  
PN W0200268649-A2.  
XX  
PD 06-SEP-2002.  
XX  
PE 31-JAN-2002; 2002WC-US002785.  
XX  
PR 31-JAN-2001; 2001US-0265395P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266767P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 07-FEB-2001; 2001US-0267057P.  
PR 08-FEB-2001; 2001US-0267459P.  
PR 09-FEB-2001; 2001US-0267823P.  
PR 15-FEB-2001; 2001US-0268974P.  
PR 26-FEB-2001; 2001US-0271664P.  
PR 27-FEB-2001; 2001US-0271839P.  
PR 27-FEB-2001; 2001US-0271855P.  
PR 02-MAR-2001; 2001US-0272788P.  
PR 02-MAR-2001; 2001US-0273046P.  
PR 14-MAR-2001; 2001US-0275925P.  
PR 14-MAR-2001; 2001US-0275947P.  
PR 14-MAR-2001; 2001US-0275950P.  
PR 14-MAR-2001; 2001US-0275989P.  
PR 15-MAR-2001; 2001US-0276448P.  
PR 15-MAR-2001; 2001US-0276450P.  
PR 16-MAR-2001; 2001US-0276397P.  
PR 16-MAR-2001; 2001US-0276768P.  
PR 20-MAR-2001; 2001US-0278652P.  
PR 26-MAR-2001; 2001US-0278775P.  
PR 26-MAR-2001; 2001US-0278778P.  
PR 29-MAR-2001; 2001US-0279882P.  
PR 29-MAR-2001; 2001US-0279884P.  
PR 30-MAR-2001; 2001US-0280147P.  
PR 11-APR-2001; 2001US-0282982P.  
PR 11-APR-2001; 2001US-0283083P.  
PR 20-APR-2001; 2001US-0285133P.  
PR 23-APR-2001; 2001US-0285749P.  
PR 03-MAY-2001; 2001US-0288327P.  
PR 03-MAY-2001; 2001US-0288504P.  
PR 29-MAY-2001; 2001US-0294047P.  
PR 30-MAY-2001; 2001US-0294473P.  
PR 08-JUN-2001; 2001US-0296964P.  
PR 18-JUN-2001; 2001US-0298959P.  
PR 19-JUN-2001; 2001US-0299324P.  
PR 13-AUG-2001; 2001US-0312020P.  
PR 16-AUG-2001; 2001US-0312889P.  
PR 16-AUG-2001; 2001US-0312988P.  
PR 21-AUG-2001; 2001US-0313309P.  
PR 28-AUG-2001; 2001US-0315470P.  
PR 31-AUG-2001; 2001US-0316447P.  
PR 07-SEP-2001; 2001US-0318115P.  
PR 07-SEP-2001; 2001US-0318118P.  
PR 12-SEP-2001; 2001US-0318740P.  
PR 19-SEP-2001; 2001US-0323379P.  
PR 18-OCT-2001; 2001US-0330245P.  
PR 18-OCT-2001; 2001US-0330308P.  
PR 14-NOV-2001; 2001US-0332701P.

XX (CURA-) CURAGEN CORP.  
PA Tchernev VT, Spylek KA, Zethusen BD, Paturajan M, Shinkets RA;  
XX Li L, Gangoli EA, Padigaru M, Anderson DW, Rastall L, Miller CE;  
PI Gelach VL, Tappier RJ, Gueev YV, Colman SD, Wolenc AR, Pena CE,  
PI Futrak K, Grosse WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;  
XX WPI; 2002-706998/76.  
DR  
XX New NOVX polypeptides and nucleic acids, useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
PS Disclosure; SEQ ID NO 420; 1498bp; English.  
XX  
CC This invention relates to a novel nucleic acids, and encoded polypeptides  
CC thereof, which have properties related to the stimulation of biochemical  
CC or physiological responses in a cell, tissue, organ or organism.  
CC Specifically, it refers to the use of biologically active fragments for  
CC diagnostic and prognostic assays and furthermore in the treatment of  
CC diverse pathological conditions. The present invention describes novel  
CC human and murine NOVX proteins, as well as methods to modulate their  
CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
CC The NOVX polypeptides, polynucleotides and antibodies are useful in  
CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,  
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
CC treating or preventing diseases such as inflammation, autoimmune  
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,  
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
CC and epilepsy. Accordingly, these molecules have many activities including  
CC cytosstatic, cardiac, antiinflammatory, immunosuppressive, antiallergic,  
CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,  
CC antiaesthetic, nephrotropic, antiarthritic, hepatotropic,  
CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,  
CC relaxant and anticonvulsant. In addition, they are useful in screening  
CC assays to identify small molecules that modulate or inhibit, for example,  
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
CC used as in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein  
CC of the invention.  
XX  
XX  
SQ Sequence 855 AA;  
Query March 99.4%; Score 4655; DB 5; Length 855;  
Best Local Similarity 99.5%; Pred. No. 1.8e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MGSDRRKGGGGPGDFGAGLKYNSRHRKVGLEGEVFLPVNNVKKYKGGPGRWVLLA 60  
Db 1 MGSDRRKGGGGPGDFGAGLKYNSRHRKVGLEGEVFLPVNNVKKYKGGPGRWVLLA 60  
QY 61 VLIIGLLVLIIGLIGFLVWHLQYRDYRVOVKXNGYRITNENFVDAYENSNSTEPLVSLASKV 120  
Db 61 VLIIGLLVLIIGLIGFLVWHLQYRDYRVOKVFGYKIRITNENFVDAYENSNSTEPLVSLASKV 120  
QY 121 KDALKLLYSVGPFLGYPYHKSAVTAFSEGSVIAYVMSBPSIPQHLVEAEKVMKEEVRVM 180  
Db 121 KDALKLLYSVGPFLGYPYHKSAVTAFSEGSVIAYVMSBPSIPQHLVEAEKVMKEEVRVM 180  
QY 181 LPPRARSILKSFVYTSVVAFPDSTKVORTQDNCSCFGLHARGVLEMTFTTGPFPDSDPYPA 240  
Db 181 LPPRARSILKSFVYTSVVAFPDSTKVORTQDNCSCFGLHARGVLEMTFTTGPFPDSDPYPA 240  
QY 241 HARGQWALRGDADSVLSLTFRSFPLASCDEKSGDLVTVYNTLSPEEHALVQLCGTTPPS 300  
Db 241 HARGQWALRGDADSVLSLTFRSFPLASCDEKSGDLVTVYNTLSPEEHALVQLCGTTPPS 300  
QY 301 YNLFHSSQWVLTITLTNTERRRHGRFEATFPOLPKMSGCGRLRKXOGTFNSPYYPGHY 360  
Db 301 YNLFHSSQWVLTITLTNTERRRHGRFEATFPOLPKMSGCGRLRKXOGTFNSPYYPGHY 360

QY 361 PENIDCTWNIENVNNOHVKVRFAKFFYLLEBPGVAGTCPKDYVEINBEKXGERSQFVYTS 420  
Db 361 PENIDCTWNIENVNNOHVKVRFAKFFYLLEBPGVAGTCPKDYVEINBEKXGERSQFVYTS 420  
QY 421 NSNKITVRHSDSYDPTGTLAETLSYSSDPFGPGFTCTGTGTCIRKELRCDGMADCTDH 480  
Db 421 NSNKITVRHSDSYDPTGTLAETLSYSSDPFGPGFTCTGTGTCIRKELRCDGMADCTDH 480  
QY 481 SDELNCSGDAGHOFTCKNKKCFKPLFWVCDSVNDGSDNSDEQSCPAQTRCNSGKCLSK 540  
Db 481 SDELNCSGDAGHOFTCKNKKCFKPLFWVCDSVNDGSDNSDEQSCPAQTRCNSGKCLSK 540  
QY 541 SQQCNKGKDCGSDSDASCPKVVVYCTKATYRCLNGELCLSKNPNBCDKEDCSGSGDEK 600  
Db 541 SQQCNKGKDCGSDSDASCPKVVVYCTKATYRCLNGELCLSKNPNBCDKEDCSGSGDEK 600  
QY 601 DCCGGLRSFTRQARVVGCTDADGEMPMQVSLHALQGHICGASLISPMVLVSAHICYID 660  
Db 601 DCCGGLRSFTRQARVVGCTDADGEMPMQVSLHALQGHICGASLISPMVLVSAHICYID 660  
QY 661 DRGFRYSDFPTQWTAFLGILDQSORSAPOVERLKRILSHPFNDFTFDYDIALLELEKP 720  
Db 661 DRGFRYSDFPTQWTAFLGILDQSORSAPOVERLKRILSHPFNDFTFDYDIALLELEKP 720  
QY 721 AEYSWVRPILCPDASHVPAGAKAIWYTGWGHYQYGGTGAIILOKGBIRVINGTCENLL 780  
Db 721 AEYSWVRPILCPDASHVPAGAKAIWYTGWGHYQYGGTGAIILOKGBIRVINGTCENLL 780  
QY 781 PQQITPRMNCVGLSGGVSCQDSSGGLPSVADGRIFGAGVSVSGDCAQSNKPGVYT 840  
Db 781 PQQITPRMNCVGLSGGVSCQDSSGGLPSVADGRIFGAGVSVSGDCAQSNKPGVYT 840  
QY 841 RLPLFRDWIKENTGV 855  
Db 841 RLPLFRDWIKENTGV 855  
RESULT 9  
AD116818  
ID AD116818 standard; protein; 855 AA.  
XX  
AC AD116818;  
XX  
DT 15-APR-2004 (first entry)  
XX  
DE Human NOVX protein homologue SegID 354.  
XX  
KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;  
KW inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; str.  
XX  
OS Homo sapiens.  
XX  
PN WO200268649-A2.  
XX  
PD 06-SEP-2002.  
XX  
PF 31-JAN-2002; 2002WC-US002785.  
XX  
PR 31-JAN-2001; 2001US-0265395P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266767P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 07-FEB-2001; 2001US-0267057P.  
PR 08-FEB-2001; 2001US-0267459P.  
PR 09-FEB-2001; 2001US-0267823P.  
PR 15-FEB-2001; 2001US-0268974P.

PR 26-FEB-2001; 2001US-0271664P.  
PR 27-FEB-2001; 2001US-0271839P.  
PR 27-FEB-2001; 2001US-0271855P.  
PR 02-MAR-2001; 2001US-0272788P.  
PR 02-MAR-2001; 2001US-0273046P.  
PR 14-MAR-2001; 2001US-0275925P.  
PR 14-MAR-2001; 2001US-0275947P.  
PR 14-MAR-2001; 2001US-0275950P.  
PR 14-MAR-2001; 2001US-0275969P.  
PR 15-MAR-2001; 2001US-0276448P.  
PR 15-MAR-2001; 2001US-0276450P.  
PR 16-MAR-2001; 2001US-0276737P.  
PR 16-MAR-2001; 2001US-0276768P.  
PR 20-MAR-2001; 2001US-0278652P.  
PR 26-MAR-2001; 2001US-0278775P.  
PR 26-MAR-2001; 2001US-0278778P.  
PR 29-MAR-2001; 2001US-0279882P.  
PR 29-MAR-2001; 2001US-0279884P.  
PR 30-MAR-2001; 2001US-0280147P.  
PR 11-APR-2001; 2001US-0282992P.  
PR 11-APR-2001; 2001US-0283083P.  
PR 20-APR-2001; 2001US-0285133P.  
PR 23-APR-2001; 2001US-0285749P.  
PR 03-MAY-2001; 2001US-0288327P.  
PR 03-MAY-2001; 2001US-0288504P.  
PR 29-MAY-2001; 2001US-0294047P.  
PR 30-MAY-2001; 2001US-0294473P.  
PR 08-JUN-2001; 2001US-0296964P.  
PR 18-JUN-2001; 2001US-0298959P.  
PR 19-JUN-2001; 2001US-0299324P.  
PR 13-AUG-2001; 2001US-0312030P.  
PR 16-AUG-2001; 2001US-0312889P.  
PR 16-AUG-2001; 2001US-0312908P.  
PR 21-AUG-2001; 2001US-0313390P.  
PR 28-AUG-2001; 2001US-0315470P.  
PR 31-AUG-2001; 2001US-0316447P.  
PR 07-SEP-2001; 2001US-0318115P.  
PR 07-SEP-2001; 2001US-0318118P.  
PR 12-SEP-2001; 2001US-0318740P.  
PR 19-SEP-2001; 2001US-0323379P.  
PR 18-OCT-2001; 2001US-0330245P.  
PR 18-OCT-2001; 2001US-0330308P.  
PR 14-NOV-2001; 2001US-0332701P.  
(CURA-) CURAGEN CORP.  
XX Tchernev VT, Spytek KA, Zernusen BD, Paturajan M, Shinketa RA;  
XX Li L, Gangoli EA, Anderson M, Pastelli L, Miller CE;  
PI Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CE;  
PI Futak K, Grose WM, Alsbrook JF, Depley DM, Rieger DK, Burgess CE;  
XX WPI; 2002-706998/76.  
DR New NOXV polypeptides and nucleic acids, useful for preventing or  
XX treating NOXV-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX Disclosure; SEQ ID NO 354; 1498bp; English.  
XX This invention relates to a novel nucleic acid, and encoded polypeptides  
XX thereof, which have properties related to the stimulation of biochemical  
XX or physiological responses in a cell, tissue, organ or organism.  
XX Specifically, it refers to the use of biologically active fragments for  
XX diagnostic and prognostic assays and furthermore in the treatment of  
XX diverse pathological conditions. The present invention describes novel  
XX human and murine NOXV proteins, as well as methods to modulate their  
XX expression using antisense oligos, ribozymes and peptide nucleic acids.  
XX The NOXV polypeptides, polynucleotides and antibodies are useful in  
XX treating or preventing NOXV-associated disorders, e.g. cardiomyopathy,  
XX atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
XX treating or preventing diseases such as inflammation, autoimmune  
XX disorders, allergies, blood disorders, acquired immunodeficiency syndrome

CC (AIDS), obesity, asthma, immunoglobulin (Ig) A nephropathy, cirrhosis,  
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
CC and epilepsy. Accordingly, these molecules have many activities including  
CC cytostatic, cardiant, antiinflammatory, immunosuppressive, antiallergic,  
CC hemostatic, anti-HIV, antidiabetic, antiatherosclerotic, anorectic,  
CC antiaesthetic, nephrotoxic, antiarthritic, hepatotoxic,  
CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,  
CC relaxant and anticonvulsant. In addition, they are useful in screening  
CC assays to identify small molecules that modulate or inhibit, for example,  
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
CC used as in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. This polypeptide is a homologue of a human NOXV protein  
XX of the invention.  
XX Sequence 855 AA:  
SQ  
Query Match 99.4%; Score 4655; DB 5; Length 855;  
Best Local Similarity 99.5%; Pred. No. 1.8e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MGSDRAKGGGKPRDPAAGLKYSRHEKXNGLEGEVFLPVNNVKKYKKGKGRWVLLAA 60  
DB 1 MGSDRAKGGGKPRDPAAGLKYSRHEKXNGLEGEVFLPVNNVKKYKKGKGRWVLLAA 60  
QY VLLGLLVLLIGIGFLVWHLQYRDVAVQKXNGYMRITNEFVDAEENSSTEFVSLASKV 120  
DB VLLGLLVLLIGIGFLVWHLQYRDVAVQKXNGYMRITNEFVDAEENSSTEFVSLASKV 120  
QY 121 KDALKLLYSGVPLGPIYHKSAAVTAPSEGSVIAIYVSEFSI PQHLYEAEARVMAEERVVM 180  
DB 121 KDALKLLYSGVPLGPIYHKSAAVTAPSEGSVIAIYVSEFSI PQHLYEAEARVMAEERVVM 180  
QY 181 LPPPARSLKSFVYTSVVAAPFTDSKTYQRTODNCSFGLHARGVLMKFTTPGPDSPYPA 240  
DB 181 LPPPARSLKSFVYTSVVAAPFTDSKTYQRTODNCSFGLHARGVLMKFTTPGPDSPYPA 240  
QY 241 HARCQNALRGDADSVLSLTFRSFDLASCDERGSDDLVTYVNTLSPMEHALVOLGTYPPS 300  
DB 241 HARCQNALRGDADSVLSLTFRSFDLASCDERGSDDLVTYVNTLSPMEHALVOLGTYPPS 300  
QY 301 YNLFHSSQNVLLITLITNTERRRHPEEATFPOLPRMSCGGRILKXQGTFSNPPYRGHY 360  
DB 301 YNLFHSSQNVLLITLITNTERRRHPEEATFPOLPRMSCGGRILKXQGTFSNPPYRGHY 360  
QY 361 PPNIDCTWNIENVNNOHVKRPKFFYLLEPGVAGTCPRKYVINGEKYCGERSQFVYTS 420  
DB 361 PPNIDCTWNIENVNNOHVKRPKFFYLLEPGVAGTCPRKYVINGEKYCGERSQFVYTS 420  
QY 421 NSNKITVRPHSDSYDTGTFLAEYLSYSSDPCEGQFTCTGRCIRKELACDGDADCTDH 480  
DB 421 NSNKITVRPHSDSYDTGTFLAEYLSYSSDPCEGQFTCTGRCIRKELACDGDADCTDH 480  
QY 481 SDEINSCDAGHOFCTCKNKECKPLFWYCDSDVNDGDNDSBQSCPAQTRFRCNGKCLSK 540  
DB 481 SDEINSCDAGHOFCTCKNKECKPLFWYCDSDVNDGDNDSBQSCPAQTRFRCNGKCLSK 540  
QY 541 SQCGNGKDDCGDSDDEASCKRNVVYTKTKTYRCINGLCLSKNPNEDGDESDGDEK 600  
DB 541 SQCGNGKDDCGDSDDEASCKRNVVYTKTKTYRCINGLCLSKNPNEDGDESDGDEK 600  
QY 601 DCDGGLSFTROARVVGAGTADGEMPMOVSLLALGQGHICGASLISPNMLVSAARCYID 660  
DB 601 DCDGGLSFTROARVVGAGTADGEMPMOVSLLALGQGHICGASLISPNMLVSAARCYID 660  
QY 661 DRGFRYSDPTQWTAFLGLHDQSORASAGVQERLKRILSHFPNDFTFDYDIALLEKRP 720  
DB 661 DRGFRYSDPTQWTAFLGLHDQSORASAGVQERLKRILSHFPNDFTFDYDIALLEKRP 720  
QY 721 AEVSWMRPCLPDASVVPFAGKAIWYTGNGHYQGTGALLIOKGIIRVINTQTCNLL 780  
DB 721 AEVSWMRPCLPDASVVPFAGKAIWYTGNGHYQGTGALLIOKGIIRVINTQTCNLL 780  
QY 781 PQGTPRMVCVGLSGVDSCQSDGSLSSVEADGRIFGAGVSWGDCGAGRNPQVYT 840

|||||  
Db 781 PQQITPRMNCVGLSGGVDSGGSGPLSSVEADGRIFQAGVYVSGDGCAGRNKPGVYT 840  
Qy 841 RLPLFRDWIKENTGV 855  
841 RLPLFRDWIKENTGV 855  
Db 841 RLPLFRDWIKENTGV 855  
RESULT 10  
AA06671  
ID AA06671 standard; protein; 855 AA.  
XX AA06671;  
XX 09-NOV-1999 (first entry)  
XX  
DE Tumour antigen derived gene-15 (TADG-15) protein.  
XX  
XX Tumour antigen derived gene-15; TADG-15; serine protease; human;  
KW breast cancer; ovary cancer; carcinoma; diagnosis.  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Domain 1..54 /note= "cytoplasmic domain"  
FT Domain 55..213 /note= "transmembrane domain"  
FT Modified-site 109..111 /note= "Asn is N-glycosylated"  
FT Region 214..447 /note= "CUB repeat"  
FT Modified-site 302..304 /note= "Asn is N-glycosylated"  
FT Region 453..602 /note= "ligand-binding repeat (class A motif)"  
FT Region 481..483 /note= "conserved SDE motif"  
FT Region 518..520 /note= "conserved SDE motif"  
FT Region 554..556 /note= "conserved SDE motif"  
FT Region 597..599 /note= "conserved SDE motif"  
FT Region /note= "conserved SDE motif"  
FT Cleavage-site 614..615  
FT Domain 615..855 /note= "catalytic domain"  
XX  
PN MO9942120-A1.  
XX  
PD 26-AUG-1999.  
XX  
PF 18-FEB-1999; 99WC-US003436.  
XX  
PR 20-FEB-1998; 98US-00027337.  
XX  
PA (UYAR-) UNIV ARKANSAS.  
XX  
PI O'Brien TJ, Tanimoto H;  
XX  
DR WPI; 1999-527418/44.  
DR N-PSDB; AAX87815.  
XX  
PT A new extracellular serine protease for diagnosis of neoplastic disease.  
XX  
PS Claim 3; Fig 10; 71pp; English.  
XX  
CC The present sequence represents a novel human extracellular serine  
CC protease, termed tumour antigen derived gene-15 protein (see AA06671),  
CC that is overexpressed in breast and ovarian carcinomas. The TADG-15 gene  
CC (see AAX87815) can be used as a diagnostic and therapeutic target in  
CC ovarian carcinoma and other carcinomas including breast, prostate, lung  
CC and colon. The ligand binding domains of TADG-15 may be valuable in the

CC uptake of specific molecules into tumour cells. The invention also  
CC provides a vector that is capable of expressing DNA encoding TADG-15  
CC protein; host cells selected from bacterial cells (especially Escherichia  
CC coli), mammalian cells, plant cells and insect cells; and a method of  
CC detecting expression of TADG-15 protein using a hybridisation probe  
XX  
SQ Sequence 855 AA;  
Query Match 99.4%; Score 4653; DB 2; Length 855;  
Best Local Similarity 99.5%; Pred. No. 2..4e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 MGSDDRARKGGGGKPDGAGLKTNSRHKVNGLEGVPLPVNNVKKYKRGGRVYLA 60  
Db 1 MGSDDRARKGGGGKPDGAGLKTNSRHKVNGLEGVPLPVNNVKKYKRGGRVYLA 60  
Qy VLTGLLVLTGLIGFLVWHLQYRDPVRVOKVNGVMRTINENFVAAYNSNSTEEVSLASKV 120  
Db VLTGLLVLTGLIGFLVWHLQYRDPVRVOKVNGVMRTINENFVAAYNSNSTEEVSLASKV 120  
Qy 121 KDALKLTVSGVPLGPHYKESAVTAESGSLVIAVYSEFSIPQHLVEAEERVAEERVVM 180  
Db 121 KDALKLTVSGVPLGPHYKESAVTAESGSLVIAVYSEFSIPQHLVEAEERVAEERVVM 180  
Qy 181 LPPRARSLSKSFVTVSVAPPTDSKTYQRTODNSCSGLHARGVELMRFTTTPGPDSPYPA 240  
Db 181 LPPRARSLSKSFVTVSVAPPTDSKTYQRTODNSCSGLHARGVELMRFTTTPGPDSPYPA 240  
Qy 241 HARCOMALRGDADSVSLTFPSFDLASCPERGDLTVVYNTLSPMEPHALVOLCGTYPPS 300  
Db 241 HARCOMALRGDADSVSLTFPSFDLASCPERGDLTVVYNTLSPMEPHALVOLCGTYPPS 300  
Qy 301 YNLTFFHSQNVLLITLITNTERRHGFEATFQLPMSSCGGRRLRAOGTFNSPYYPGHY 360  
Db 301 YNLTFFHSQNVLLITLITNTERRHGFEATFQLPMSSCGGRRLRAOGTFNSPYYPGHY 360  
Qy 361 PENIDCTWNIENVNNOHKYRFFEFYLLBPGVAGTCPKDYVINEBKTCGEGSQPVVYS 420  
Db 361 PENIDCTWNIENVNNOHKYRFFEFYLLBPGVAGTCPKDYVINEBKTCGEGSQPVVYS 420  
Qy 421 NSNKITVRFHSDSYDTDTGFLAETLSYDSSDPGQFTCRGTGCIKRELRCDMACTDH 480  
Db 421 NSNKITVRFHSDSYDTDTGFLAETLSYDSSDPGQFTCRGTGCIKRELRCDMACTDH 480  
Qy 481 SDELINSCDAGHOFCKNKECKPLFWVCDSVNDGDNSSDEGSCCPAQTFRCNGKCLSK 540  
Db 481 SDELINSCDAGHOFCKNKECKPLFWVCDSVNDGDNSSDEGSCCPAQTFRCNGKCLSK 540  
Qy 541 SQCCNGKDCGSDGSDASCPKXNVVCTKRTYCLNGLCLSKNPPCDGKEDSDSDSK 600  
Db 541 SQCCNGKDCGSDGSDASCPKXNVVCTKRTYCLNGLCLSKNPPCDGKEDSDSDSK 600  
Qy 601 DCCGGLRSFTFRQARVVGTDADGEMPMOVSLLALQGHICGASLISPMVLVSAACIYD 660  
Db 601 DCCGGLRSFTFRQARVVGTDADGEMPMOVSLLALQGHICGASLISPMVLVSAACIYD 660  
Qy 661 DRGFRYSDFPTQWTAFLGLHDQSORSAFGVERLTKIISHPFNDFTFDYDIALLEKXP 720  
Db 661 DRGFRYSDFPTQWTAFLGLHDQSORSAFGVERLTKIISHPFNDFTFDYDIALLEKXP 720  
Qy 721 AAYSSMWRPCLPDAHVPPAGKAIWVTGWGHTQYGGTALLIQKEIRIYNTTBENLL 780  
Db 721 AAYSSMWRPCLPDAHVPPAGKAIWVTGWGHTQYGGTALLIQKEIRIYNTTBENLL 780  
Qy 781 PQQITPRMNCVGLSGGVDSGGSGPLSSVEADGRIFQAGVYVSGDGCAGRNKPGVYT 840  
Db 781 PQQITPRMNCVGLSGGVDSGGSGPLSSVEADGRIFQAGVYVSGDGCAGRNKPGVYT 840  
Qy 841 RLPLFRDWIKENTGV 855  
Db 841 RLPLFRDWIKENTGV 855



RESULT 11  
 AAB98500  
 ID AAB98500 standard; protein; 855 AA.  
 XX  
 AC AAB98500;  
 XX  
 DT 03-AUG-2001 (first entry)  
 XX  
 DE Human TADG-15.  
 XX  
 KW Human; TADG-15; cytosolic; vaccine; ovarian tumour; cancer;  
 KW tumour antigen-derived gene 15; extracellular serine protease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200129056-A1.  
 XX  
 PD 26-APR-2001.  
 XX  
 PF 20-OCT-2000; 2000MO-US029095.  
 XX  
 PR 20-OCT-1999; 99US-00421213.  
 XX  
 PA (UYAR-) UNIV ARKANSAS.  
 XX  
 PI O'Brien TJ, Tanihato H;  
 XX  
 DR WPI; 2001-381031/40.  
 DR N-PSDB; AAH23601.  
 XX  
 PT Novel extracellular serine protease, termed tumor antigen-derived gene 15  
 PT protein overexpressed in carcinomas and DNA encoding it, for diagnosis,  
 PT treatment, prevention of cancer, particularly breast, ovarian cancer.  
 XX  
 PS Claim 11; Fig 2; 130pp; English.  
 XX  
 CC The present sequence represents human tumour antigen-derived gene 15  
 CC (TADG-15) protein. TADG-15 is an extracellular serine protease. It was  
 CC found that TADG-15 is over-expressed in ovarian tumours. TADG-15 protein  
 CC or its fragments of 9-20 residues that lack TADG-15 protease activity are  
 CC useful for vaccinating an individual against TADG-15, having, suspected  
 CC of having or at risk of getting cancer. Furthermore, the TADG-15 gene can  
 CC be used as a diagnostic or therapeutic target in cancer  
 CC  
 SQ Sequence 855 AA;  
 Query Match 99.4%; Score 4653; DB 4; Length 855;  
 Best Local Similarity 99.5%; Pred. No. 2.4e-307;  
 Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MGSRRARRGGGGPDPFGAGLKYNSRHEKVNGLSEGEVEFLPVNNYKVKYKGPGRWVTLAA 60  
 DB 1 MGSRRARRGGGGPDPFGAGLKYNSRHEKVNGLSEGEVEFLPVNNYKVKYKGPGRWVTLAA 60  
 QY 61 VLIGLLVLIGLIGFLVHMLQYRDYRVCKYKNGYRITENFVDAYENSNSTEFLVSLASKV 120  
 DB 61 VLIGLLVLIGLIGFLVHMLQYRDYRVCKYKNGYRITENFVDAYENSNSTEFLVSLASKV 120  
 QY 121 KDALKLYSGVPFLGPKYHESAVTAFSEGSVIAYYMSFSLPQHLVEAEERVMABEERVVM 180  
 DB 121 KDALKLYSGVPFLGPKYHESAVTAFSEGSVIAYYMSFSLPQHLVEAEERVMABEERVVM 180  
 QY 121 LPPPARSKSFVWVSVAFPPTDSKTVOQTQNSCSFGHLAAGVLELMRTTGGPDSPPPA 240  
 DB 121 LPPPARSKSFVWVSVAFPPTDSKTVOQTQNSCSFGHLAAGVLELMRTTGGPDSPPPA 240  
 QY 181 LPPPARSKSFVWVSVAFPPTDSKTVOQTQNSCSFGHLAAGVLELMRTTGGPDSPPPA 240  
 DB 181 LPPPARSKSFVWVSVAFPPTDSKTVOQTQNSCSFGHLAAGVLELMRTTGGPDSPPPA 240  
 QY 241 HARCQMALRGDADSVLSTFRSPDLASCDERGSDELVTYNTLSPMEPALVOLGCTYPPS 300  
 DB 241 HARCQMALRGDADSVLSTFRSPDLASCDERGSDELVTYNTLSPMEPALVOLGCTYPPS 300  
 QY 301 YNLTFHSSQNVLLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTNSPYPGHY 360  
 DB 301 YNLTFHSSQNVLLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTNSPYPGHY 360

QY 361 PENIDCTWNIEVPNNQHVKVFKEFFYLLEPGVAGTCPKDVEINGEKYCGERSQFVVT 420  
 DB 361 PENIDCTWNIEVPNNQHVKVSFKFFYLLEPGVAPGTCPKQVEINGEKYCGERSQFVVT 420  
 QY 421 NSNKITVRPHSDQSYDTGTGLAEYLSYDSSDPCPGQFCRTGRCIRKELRCDGNADCTDH 480  
 DB 421 NSNKITVRPHSDQSYDTGTGLAEYLSYDSSDPCPGQFCRTGRCIRKELRCDGNADCTDH 480  
 QY 481 SDELNCSGDAGHQTCKNRCCKPLFWVCDSVNDGDNSSDQSCSPQOTRCSNGKCLSK 540  
 DB 481 SDELNCSGDAGHQTCKNRCCKPLFWVCDSVNDGDNSSDQSCSPQOTRCSNGKCLSK 540  
 QY 541 SQQNGKDDCGDSDSEASCPVNVVCTCKHYRCLANGCLCSKNGPEDCKEDCSGSDDEK 600  
 DB 541 SQQNGKDDCGDSDSEASCPVNVVCTCKHYRCLANGCLCSKNGPEDCKEDCSGSDDEK 600  
 QY 601 DCDGGLRSFTRQARVVGTDADBEQWPMQVSLHALGGGHCASLISPNMLVSAHACYID 660  
 DB 601 DCDGGLRSFTRQARVVGTDADBEQWPMQVSLHALGGGHCASLISPNMLVSAHACYID 660  
 QY 661 DRGFRYSDPTQWTFELGLHQDSQASAPGVQERLKRILISHPFNDFTFDYDIALLELEKP 720  
 DB 661 DRGFRYSDPTQWTFELGLHQDSQASAPGVQERLKRILISHPFNDFTFDYDIALLELEKP 720  
 QY 721 AEYSMWRTPICLPDASHVPFPAKAIWVTGNGHTQYGGTGALILQKGIIRVYNOTTCEML 780  
 DB 721 AEYSMWRTPICLPDASHVPFPAKAIWVTGNGHTQYGGTGALILQKGIIRVYNOTTCEML 780  
 QY 781 PQQITPRMNCVGLSGVDSCQDGSGLSSVEADGRIFGAGVSWGDGCAGRNKPQVYT 840  
 DB 781 PQQITPRMNCVGLSGVDSCQDGSGLSSVEADGRIFGAGVSWGDGCAGRNKPQVYT 840  
 QY 841 RLPLFRDMIKENTGV 855  
 DB 841 RLPLFRDMIKENTGV 855  
 RESULT 12  
 AAB06930  
 ID AAB06930 standard; protein; 855 AA.  
 XX  
 AC AAB06930;  
 XX  
 DT 16-OCT-2001 (first entry)  
 XX  
 DE Human membrane-type serine protease (MTSP) 1.  
 XX  
 KW Human; transmembrane serine protease; membrane-type serine protease;  
 KW MTSP; protease domain; neoplastic disease; tumour; cancer; cytosolic;  
 KW lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;  
 KW matrix.ase.  
 XX  
 OS Homo sapiens.  
 XX  
 PN Key location/Qualifiers  
 FT Domain 615..855  
 FT /label= Protease\_domain  
 MO200157194-A2.  
 PD 09-AUG-2001.  
 PD 02-FEB-2001; 2001MO-US003471.  
 XX  
 PR 03-FEB-2000; 2000US-0179982P.  
 PR 18-FEB-2000; 2000US-0183542P.  
 PR 22-JUN-2000; 2000US-0231342P.  
 PR 26-JUL-2000; 2000US-0230970P.  
 PR 08-SEP-2000; 2000US-0065798P.  
 PR 22-SEP-2000; 2000US-0234840P.  
 XX  
 PA (CORV-) CORVAS INT INC.

XX Madison EL, Ong EO, Yeh J;  
 XX WPI; 2001-468877/53.  
 DR N-PSDB; AAD13113.  
 XX Novel single chain polypeptide comprising protease domain of type-II  
 PT membrane-type serine protease or its catalytically active portion useful  
 PT for treating and preventing cancer and tumor.  
 XX  
 PS Claim 12; Page 195-197; 256pp; English.  
 XX  
 CC The invention relates to transmembrane serine proteases and their  
 CC corresponding nucleotides and the protease domain of a type-II membrane-  
 CC type serine protease (MTSP). MTSP is useful for identifying compounds  
 CC that modulate or inhibit its proteolytic activity and for formulating a  
 CC medicament for treating neoplastic disease. MTSP and its corresponding  
 CC nucleotides are useful in preventing or treating tumors or cancers such  
 CC as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in  
 CC diagnostics and in hybridisation assays. MTSP is useful as a diagnostic  
 CC marker for tumour development, growth and/or progression and as  
 CC immunogens to generate antibodies that specifically bind to it. MTSP DNA  
 CC is useful in a yeast two-hybrid system and in gene therapy. The present  
 CC sequence is human MTSP1 protein (also called matrilipase)  
 CC  
 XX Sequence 855 AA;  
 SQ  
 Query Match 99.4%; Score 4653; DB 4; Length 855;  
 Best Local Similarity 99.5%; Pred. No. 2.4e-307;  
 Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MGSDDRARKGGGPGKFGAGLKYNSRHEKVNGLBEVFLPVNNVKKYKIGPGRVVLAA 60  
 DB 1 MGSDDRARKGGGPGKFGAGLKYNSRHEKVNGLBEVFLPVNNVKKYKIGPGRVVLAA 60  
 QY 1 VLGILLVLTIGITVHMLQYRDVRYOKNKGYNRINENPVDAYENSNTSEPSLASKV 120  
 DB 61 VLGILLVLTIGITVHMLQYRDVRYOKNKGYNRINENPVDAYENSNTSEPSLASKV 120  
 QY 121 KDALKLLYSGVPFPGPHKESAVTAFSEGSVIAYYSEFSIPGLVVEARVMAEERVM 180  
 DB 121 KDALKLLYSGVPFPGPHKESAVTAFSEGSVIAYYSEFSIPGLVVEARVMAEERVM 180  
 QY 181 LPPRARSLSKFPVYTSVVAFPDTSKTVORTDNGSCFGLHARGVELMFTTGPFPDSPA 240  
 DB 181 LPPRARSLSKFPVYTSVVAFPDTSKTVORTDNGSCFGLHARGVELMFTTGPFPDSPA 240  
 QY 241 HARCQWALRGDADSVSLTFRSPDLASCDRGSILYVNTLSMEPHALVOLCGTYPSP 300  
 DB 241 HARCQWALRGDADSVSLTFRSPDLASCDRGSILYVNTLSMEPHALVOLCGTYPSP 300  
 QY 301 YNLTFRSSQVNLITLTITNTERRHPPGEATFPOLPRMSSCGRLRKQGTFNSPYYPGHY 360  
 DB 301 YNLTFRSSQVNLITLTITNTERRHPPGEATFPOLPRMSSCGRLRKQGTFNSPYYPGHY 360  
 QY 361 PENIDCTWNIETVNNQHVAKRFKFFYLLEPGVPAAGTCPKDYVEINGEKYCGERSQFVTS 420  
 DB 361 PENIDCTWNIETVNNQHVAKRFKFFYLLEPGVPAAGTCPKDYVEINGEKYCGERSQFVTS 420  
 QY 421 NSKKITRFRHSDDSYDTGTGLAEYLSVSDSDPCGQGTCTGTCIRKELACDGMADCTDH 480  
 DB 421 NSKKITRFRHSDDSYDTGTGLAEYLSVSDSDPCGQGTCTGTCIRKELACDGMADCTDH 480  
 QY 481 SDEINSCDAGHOPCTGNKCKCPLEFWICDSVNDGSDSDGSCSPAQTRFCSNGKCLSK 540  
 DB 481 SDEINSCDAGHOPCTGNKCKCPLEFWICDSVNDGSDSDGSCSPAQTRFCSNGKCLSK 540  
 QY 541 SQQCNKGKDCGSDGSDASCPKVVVVTCTKTYRCLNGLCLSKNPECDCGKEDCSGSDSK 600  
 DB 541 SQQCNKGKDCGSDGSDASCPKVVVVTCTKTYRCLNGLCLSKNPECDCGKEDCSGSDSK 600  
 QY 601 DCCGGLRSFTROARVVGCTADBEEMPWQVSHALGQGHICGASLISPMWLVSAHACYID 660  
 DB 601 DCCGGLRSFTROARVVGCTADBEEMPWQVSHALGQGHICGASLISPMWLVSAHACYID 660

DB 601 DCCGGLRSFTROARVVGCTADBEEMPWQVSHALGQGHICGASLISPMWLVSAHACYID 660  
 QY 661 DRGFRYSDPQTWTAFLGLHDQSORSAPOVERLKRILISHPFNDFTFDYDIALLELEXP 720  
 DB 661 DRGFRYSDPQTWTAFLGLHDQSORSAPOVERLKRILISHPFNDFTFDYDIALLELEXP 720  
 QY 721 AEYSWVRPILCPDASHVPFAGKAIWVTGHTQYGGTGLILQKEIRVYNQTTBENLL 780  
 DB 721 AEYSWVRPILCPDASHVPFAGKAIWVTGHTQYGGTGLILQKEIRVYNQTTBENLL 780  
 QY 781 PQQITRRMVCVGLSGGVSCGDSGGLSSVSEADRIIGAGVYVSWGDCAGRNKQCVYT 840  
 DB 781 PQQITRRMVCVGLSGGVSCGDSGGLSSVSEADRIIGAGVYVSWGDCAGRNKQCVYT 840  
 QY 841 RLPLFRDWIKENTGV 855  
 DB 841 RLPLFRDWIKENTGV 855  
 RESULT 13  
 AAO22929  
 ID AAO22929 standard; protein; 855 AA.  
 AC AAO22929;  
 XX  
 XX 12-DEC-2002 (first entry)  
 DE Type II transmembrane serine protease 1 protein SEQ ID No 2.  
 XX  
 KW Cytostatic; type-II membrane-type serine protease 7; MTSP7; malignancy;  
 KW neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;  
 OS malignant; enzyme.  
 XX Homo sapiens.  
 XX  
 PN WO200272786-A2.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 13-MAR-2002; 2002WO-US007903.  
 XX  
 PR 13-MAR-2001; 2001US-0275592P.  
 XX  
 PA (CORV-) CORVAS INT INC.  
 XX  
 PI Madison EL, Ong EO;  
 XX  
 DR WPI; 2002-732827/79.  
 DR N-PSDB; AAL53444.  
 XX  
 PT New transmembrane serine protease 7 (MTSP7) polypeptide for diagnosing  
 PT neoplastic diseases, monitoring tumor progress or therapeutic  
 PT effectiveness, or identifying MTSP7 modulators for treating tumors or  
 PT cancers.  
 XX  
 PS Disclosure; Page 172-174; 184pp; English.  
 XX  
 CC The invention relates to a purified single or two-chain polypeptide,  
 CC which comprises the protease domain of a type-II membrane-type serine  
 CC protease 7 (MTSP7) or its catalytically active portion. The polypeptide  
 CC comprising MTSP7 is useful for detecting or diagnosing a neoplastic  
 CC disease, a pre-malignant lesion, a malignancy or other pathological  
 CC condition in a subject. This polypeptide is also useful for monitoring  
 CC tumor (e.g. tumor of the breast, cervix, prostate, lung, ovary or  
 CC colon) progress and/or therapeutic effectiveness. The inhibitor of the  
 CC polypeptide or the inhibitor of the polypeptide's zymogen form is useful  
 CC for treating or preventing a neoplastic disease, or tumor initiation,  
 CC growth or progression, or a (pre-)malignant condition. The polypeptide or  
 CC polynucleotide is also useful for identifying modulators of MTSP7, which  
 CC may be used to treat cancers or tumors. This sequence represents a  
 CC protein of the type-II membrane-type serine protease 1 relating to the  
 CC invention  
 XX

SQ Sequence 855 AA;

Query Match 99.4%; Score 4653; DB 5; Length 855;  
Best Local Similarity 99.5%; Pred. No. 2,4e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDRAKGGGGPDPFAGLKYNSRHEKYNLEGEVEFLPVNNYKVKYKPGHVVVLA 60  
DB 1 MGSDRAKGGGGPDPFAGLKYNSRHEKYNLEGEVEFLPVNNYKVKYKPGHVVVLA 60

QY 61 VLIGLLVLIGLIGFLVHLYQYRDYRVQKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 120  
DB 61 VLIGLLVLIGLIGFLVHLYQYRDYRVQKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 120

QY 121 KDALKLYSGVPFLGPHYKESAVTAFSGSIYATYMSFSIYPOHLVEAEVMAEERVVM 180  
DB 121 KDALKLYSGVPFLGPHYKESAVTAFSGSIYATYMSFSIYPOHLVEAEVMAEERVVM 180

QY 181 LPPARSLKSFVYVSVAFFPDSKTQRTQDSCSFGIHAAGVEMRPTTGPPDSFYPA 240  
DB 181 LPPARSLKSFVYVSVAFFPDSKTQRTQDSCSFGIHAAGVEMRPTTGPPDSFYPA 240

QY 241 HARCQMLRGDADSVLITFRSFDLASCDEGSDLVTVYNTLSPEMHPALVQLGCTYPPS 300  
DB 241 HARCQMLRGDADSVLITFRSFDLASCDEGSDLVTVYNTLSPEMHPALVQLGCTYPPS 300

QY 301 YNLTFHSSQNVLLTTLTNTERRHGFATFPOLPRMSSCGGRIRKAOGTNSPYRHY 360  
DB 301 YNLTFHSSQNVLLTTLTNTERRHGFATFPOLPRMSSCGGRIRKAOGTNSPYRHY 360

QY 361 PPNIDCTWNIEVPNNQHKVAFKFFYLLEPGVATGCKDVEYNGEKYCGERQFOVYVS 420  
DB 361 PPNIDCTWNIEVPNNQHKVAFKFFYLLEPGVATGCKDVEYNGEKYCGERQFOVYVS 420

QY 421 NSNKTIVRFHSDQYTDGTFLAELYSYSDSPCEQFTCRTGRCIRKELRCGMACTDH 480  
DB 421 NSNKTIVRFHSDQYTDGTFLAELYSYSDSPCEQFTCRTGRCIRKELRCGMACTDH 480

QY 481 SDELNCGDAGHOTCCRRKPKPLFWVCDSYNDGDNDSDEGSCCPAOTPRCSNGKCLSK 540  
DB 481 SDELNCGDAGHOTCCRRKPKPLFWVCDSYNDGDNDSDEGSCCPAOTPRCSNGKCLSK 540

QY 541 SQQNGKDDCGDSDSCPKVNVYTCMKHYRCCLNGCLSKNPECDCGKEDCGSDSEK 600  
DB 541 SQQNGKDDCGDSDSCPKVNVYTCMKHYRCCLNGCLSKNPECDCGKEDCGSDSEK 600

QY 601 DCDGGLRSFTRQARVVGTDADBEQWQVSLHALGQGHICGASLISPNNLVSAHACYID 660  
DB 601 DCDGGLRSFTRQARVVGTDADBEQWQVSLHALGQGHICGASLISPNNLVSAHACYID 660

QY 661 DRGRYSDPTQWTFGLIHDSQSRAPGVQRRRLKRIISHPFNDFTFYDIALLELKP 720  
DB 661 DRGRYSDPTQWTFGLIHDSQSRAPGVQRRRLKRIISHPFNDFTFYDIALLELKP 720

QY 721 AEYSWVRPILPASHVFPAGKAIWYMGHTQGGTIGALLKGBELRVNTQTCLELL 780  
DB 721 AEYSWVRPILPASHVFPAGKAIWYMGHTQGGTIGALLKGBELRVNTQTCLELL 780

QY 781 PQQITPRMWCVFLSGVDSCQDSCGFLSVEADGRIFGAGVSWGDCAGRRKPGVYT 840  
DB 781 PQQITPRMWCVFLSGVDSCQDSCGFLSVEADGRIFGAGVSWGDCAGRRKPGVYT 840

QY 841 RLPLFRDWIKENTGV 855  
DB 841 RLPLFRDWIKENTGV 855

RESULT 14  
AD116816  
ID AD116816 standard; protein, 855 AA.  
XX AC AD116816;  
XX

DT 15-APR-2004 (first entry)  
XX  
DE Human NOX protein homologue Segid 352.  
XX human; NOX; cardiomyopathy; atherosclerosis; cancer; diabetes;  
KW inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig) A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; str.  
XX  
OS Homo sapiens.  
XX  
PN W020268649-A2.  
XX  
PD 06-SEP-2002.  
XX  
PF 31-JAN-2002; 2002WC-US002785.  
XX  
PR 31-JAN-2001; 2001US-0265395P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266767P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 07-FEB-2001; 2001US-0267057P.  
PR 08-FEB-2001; 2001US-0267459P.  
PR 09-FEB-2001; 2001US-0267823P.  
PR 15-FEB-2001; 2001US-0268974P.  
PR 26-FEB-2001; 2001US-0271664P.  
PR 27-FEB-2001; 2001US-0271839P.  
PR 27-FEB-2001; 2001US-0271855P.  
PR 02-MAR-2001; 2001US-0272788P.  
PR 02-MAR-2001; 2001US-0273046P.  
PR 14-MAR-2001; 2001US-0275925P.  
PR 14-MAR-2001; 2001US-0275947P.  
PR 14-MAR-2001; 2001US-0275950P.  
PR 14-MAR-2001; 2001US-0275989P.  
PR 15-MAR-2001; 2001US-0276448P.  
PR 15-MAR-2001; 2001US-0276450P.  
PR 16-MAR-2001; 2001US-0276397P.  
PR 16-MAR-2001; 2001US-0276768P.  
PR 20-MAR-2001; 2001US-0278652P.  
PR 26-MAR-2001; 2001US-0278775P.  
PR 26-MAR-2001; 2001US-0278778P.  
PR 29-MAR-2001; 2001US-0279882P.  
PR 29-MAR-2001; 2001US-0279884P.  
PR 30-MAR-2001; 2001US-0280147P.  
PR 11-APR-2001; 2001US-0282982P.  
PR 11-APR-2001; 2001US-0283083P.  
PR 20-APR-2001; 2001US-0285133P.  
PR 23-APR-2001; 2001US-0285749P.  
PR 03-MAY-2001; 2001US-0288377P.  
PR 03-MAY-2001; 2001US-0288504P.  
PR 29-MAY-2001; 2001US-0294047P.  
PR 30-MAY-2001; 2001US-0294473P.  
PR 08-JUN-2001; 2001US-0296564P.  
PR 18-JUN-2001; 2001US-0298959P.  
PR 19-JUN-2001; 2001US-0299349P.  
PR 13-AUG-2001; 2001US-0312020P.  
PR 16-AUG-2001; 2001US-0312889P.  
PR 16-AUG-2001; 2001US-0312968P.  
PR 21-AUG-2001; 2001US-0313390P.  
PR 28-AUG-2001; 2001US-0315470P.  
PR 31-AUG-2001; 2001US-0316447P.  
PR 07-SEP-2001; 2001US-0318115P.  
PR 12-SEP-2001; 2001US-0318118P.  
PR 12-SEP-2001; 2001US-0318740P.  
PR 19-SEP-2001; 2001US-0323379P.  
PR 18-OCT-2001; 2001US-0330245P.  
PR 18-OCT-2001; 2001US-0330308P.  
PR 14-NOV-2001; 2001US-0332701P.  
XX

PA (CURA-) CURAGEN CORP.  
XX Therneev VT, Splytek KA, Zernusen BD, Paturajan M, Shinkera RA;  
PI Li L, Ganggoli EA, Padigan M, Anderson SD, Rastelli L, Miller CE;  
PI Gerlach VL, Taupier RO, Gueev VY, Colman DW, Wolenc AR, Pena CB;  
PI Futrak K, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK, Burgess CE;  
XX WPI; 2002-706998/76.  
XX New NOVX polypeptides and nucleic acids, useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
PS Disclosure; SEQ ID NO 352; 1498bp; English.  
XX  
XX This invention relates to a novel nucleic acids, and encoded polypeptides  
CC thereof, which have properties related to the stimulation of biochemical  
CC or physiological responses in a cell, tissue, organ or organism.  
CC Specifically, it refers to the use of biologically active fragments for  
CC diagnostic and prognostic assays and furthermore in the treatment of  
CC diverse pathological conditions. The present invention describes novel  
CC human and murine NOVX proteins, as well as methods to modulate their  
CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
CC The NOVX polypeptides, polynucleotides and antibodies are useful in  
CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,  
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
CC treating or preventing diseases such as inflammation, autoimmune  
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,  
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
CC and epilepsy. Accordingly, these molecules have many activities including  
CC cytosstatic, cardiac, anti-inflammatory, immunosuppressive, antiallergic,  
CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,  
CC antiaesthetic, nephroprotective, antiarthritic, hepatotropic,  
CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,  
CC relaxant and antiviral. In addition, they are useful in screening  
CC assays to identify small molecules that modulate or inhibit, for example,  
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
CC used as in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein  
CC of the invention.  
XX  
XX Sequence 855 AA;  
SQ  
Query Match 99.4%; Score 4653; DB 5; Length 855;  
Best Local Similarity 99.5%; Pred. No. 2.4e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MGSDRARKGGGPGKDFGAGLKYNSRHRKVNGLBEGVEFLPVNNVKKYKKGPGRWVTLAA 60  
DB 1 MGSDRARKGGGPGKDFGAGLKYNSRHRKVNGLBEGVEFLPVNNVKKYKKGPGRWVTLAA 60  
QY 61 VLIGLLVLLIGLIGLVNHLQRYDRVQVKNKGYKRTINEFVDAYENSNSTEPIASLKV 120  
DB 61 VLIGLLVLLIGLIGLVNHLQRYDRVQVKNKGYKRTINEFVDAYENSNSTEPIASLKV 120  
QY 121 KDALKLKYSGVPLGPGYHKSAVTAFFSEGVIATYMSSEFSIPQHLVEAEAEVMAEERVVM 180  
DB 121 KDALKLKYSGVPLGPGYHKSAVTAFFSEGVIATYMSSEFSIPQHLVEAEAEVMAEERVVM 180  
QY 181 LPPRARSILKSFVVTSSVVAFTDSKTIVRTQDNCSCFGLHARGVLMFTTGPFPDSYPYA 240  
DB 181 LPPRARSILKSFVVTSSVVAFTDSKTIVRTQDNCSCFGLHARGVLMFTTGPFPDSYPYA 240  
QY 241 HARQVRLRGDADSVLSITRSPFLASCDEBGSPLVTVNTLSMEBHALVOLCGTTPPS 300  
DB 241 HARQVRLRGDADSVLSITRSPFLASCDEBGSPLVTVNTLSMEBHALVOLCGTTPPS 300  
QY 301 YNLFEHSSONVLTITLTNTERRRHPGEATFPOLPRMSSCGGRLRKAOGTFNSPYYPGHY 360  
DB 301 YNLFEHSSONVLTITLTNTERRRHPGEATFPOLPRMSSCGGRLRKAOGTFNSPYYPGHY 360

QY 361 PENIDCTWNIENVNNOHVKVRFEKFFYLLEPGVAGTCPODYVINEKRYCGERSQFVVT 420  
DB 361 PENIDCTWNIENVNNOHVKVSFKFFYLLEPGVAGTCPODYVINEKRYCGERSQFVVT 420  
QY 421 NSKRTVRRHSDSYNDTGTFLAEYLSYDSSDPGQPTCTGTGCTCKELRCQWMACTTH 480  
DB 421 NSKRTVRRHSDSYNDTGTFLAEYLSYDSSDPGQPTCTGTGCTCKELRCQWMACTTH 480  
QY 481 SDELNCSDAQHOFCTCKNFKCKPLFWVCDSVNDCCGNSDEQSCCPAQTFRCNGKCLSK 540  
DB 481 SDELNCSDAQHOFCTCKNFKCKPLFWVCDSVNDCCGNSDEQSCCPAQTFRCNGKCLSK 540  
QY 541 SQQCNKGKDCGSDASCPKVVNTCTYKTRTCLNGLCLSKNPECDGKEDCSGSDSK 600  
DB 541 SQQCNKGKDCGSDASCPKVVNTCTYKTRTCLNGLCLSKNPECDGKEDCSGSDSK 600  
QY 601 DCCCGLRSPTRQARVVGCTDADGEMPMQVSLHALGQGHICGASLSPMWLSAAHCYTD 660  
DB 601 DCCCGLRSPTRQARVVGCTDADGEMPMQVSLHALGQGHICGASLSPMWLSAAHCYTD 660  
QY 661 DRGFRYSDFPQMTAFGLHDQSORSAPQOERLKRILSHPFENDTFPYDIALLEKRP 720  
DB 661 DRGFRYSDFPQMTAFGLHDQSORSAPQOERLKRILSHPFENDTFPYDIALLEKRP 720  
QY 721 AEYSSWVRPICLPDASHVPAGKAIWVTGSHGTQYGTGALLIQQEIRVINOCTCENLL 780  
DB 721 AEYSSWVRPICLPDASHVPAGKAIWVTGSHGTQYGTGALLIQQEIRVINOCTCENLL 780  
QY 781 PQGTRRMVCVGLSGVNDSCGDSGGLSVYADRIIGAGVSVSGDCAGNKRKYVT 840  
DB 781 PQGTRRMVCVGLSGVNDSCGDSGGLSVYADRIIGAGVSVSGDCAGNKRKYVT 840  
QY 841 RLPLFRDWIKENTGV 855  
DB 841 RLPLFRDWIKENTGV 855  
RESULT 15  
AD116882  
ID AD116882 standard; protein; 855 AA.  
XX  
XX AD116882;  
DT 15-APR-2004 (first entry)  
XX  
XX Human NOVX protein homologue Segid 418.  
DE  
XX human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;  
KW inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; str.  
XX  
XX Homo sapiens.  
PN  
XX NC0200268649-A2.  
PD  
XX 06-SEP-2002.  
PF  
XX 31-JAN-2002; 2002MO-US002785.  
XX  
XX  
XX 31-JAN-2001; 2001US-0265395P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266767P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 08-FEB-2001; 2001US-0267057P.  
PR 09-FEB-2001; 2001US-0267459P.  
PR 15-FEB-2001; 2001US-0268742P.  
PR 26-FEB-2001; 2001US-0271664P.

PR 27-FEB-2001; 2001US-0271839P.  
 PR 27-FEB-2001; 2001US-0271855P.  
 PR 02-MAR-2001; 2001US-0272788P.  
 PR 02-MAR-2001; 2001US-0273046P.  
 PR 14-MAR-2001; 2001US-0275925P.  
 PR 14-MAR-2001; 2001US-0275947P.  
 PR 14-MAR-2001; 2001US-0275950P.  
 PR 14-MAR-2001; 2001US-0275989P.  
 PR 15-MAR-2001; 2001US-0276448P.  
 PR 15-MAR-2001; 2001US-0276450P.  
 PR 16-MAR-2001; 2001US-0276397P.  
 PR 16-MAR-2001; 2001US-0276768P.  
 PR 20-MAR-2001; 2001US-0278652P.  
 PR 26-MAR-2001; 2001US-0278775P.  
 PR 26-MAR-2001; 2001US-0278778P.  
 PR 29-MAR-2001; 2001US-0279882P.  
 PR 29-MAR-2001; 2001US-0279884P.  
 PR 30-MAR-2001; 2001US-0280147P.  
 PR 11-APR-2001; 2001US-0282929P.  
 PR 11-APR-2001; 2001US-0283083P.  
 PR 20-APR-2001; 2001US-0285133P.  
 PR 23-APR-2001; 2001US-0285749P.  
 PR 03-MAY-2001; 2001US-0288327P.  
 PR 03-MAY-2001; 2001US-0288504P.  
 PR 29-MAY-2001; 2001US-0294047P.  
 PR 30-MAY-2001; 2001US-0294473P.  
 PR 08-JUN-2001; 2001US-0296964P.  
 PR 18-JUN-2001; 2001US-0298959P.  
 PR 19-JUN-2001; 2001US-0299324P.  
 PR 13-AUG-2001; 2001US-0312020P.  
 PR 16-AUG-2001; 2001US-0312889P.  
 PR 16-AUG-2001; 2001US-0312908P.  
 PR 21-AUG-2001; 2001US-0313390P.  
 PR 28-AUG-2001; 2001US-0315470P.  
 PR 31-AUG-2001; 2001US-0316447P.  
 PR 07-SEP-2001; 2001US-0318115P.  
 PR 07-SEP-2001; 2001US-0318118P.  
 PR 12-SEP-2001; 2001US-0318740P.  
 PR 19-SEP-2001; 2001US-0323379P.  
 PR 18-OCT-2001; 2001US-0330245P.  
 PR 18-OCT-2001; 2001US-0330308P.  
 PR 14-NOV-2001; 2001US-0332701P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 XX Tchernev VT, Spytek KA, Zernusen BD, Patturajan M, Shimkets RA;  
 PI Li L, Gangoli EA, Padigaru M, Anderson DW, Rastelli L, Miller CE;  
 PI Gerlach VT, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CEA;  
 PI Futrak K, Grose WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;  
 XX  
 DR WPI; 2002-706998/76.  
 XX  
 XX New NOXV polypeptides and nucleic acids, useful for preventing or  
 PT treating NOXV-associated disorders, e.g. cancer, cardiomyopathy,  
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 PT  
 XX  
 PS Disclosure; SEQ ID NO 418; 1498bp; English.  
 XX  
 CC This invention relates to a novel nucleic acid, and encoded polypeptides  
 CC thereof, which have properties related to the stimulation of biochemical  
 CC or physiological responses in a cell, tissue, organ or organism.  
 CC Specifically, it refers to the use of biologically active fragments for  
 CC diagnostic and prognostic assays and furthermore in the treatment of  
 CC diverse pathological conditions. The present invention describes novel  
 CC human and murine NOXV proteins, as well as methods to modulate their  
 CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
 CC The NOXV polypeptides, polynucleotides and antibodies are useful in  
 CC treating or preventing NOXV-associated disorders, e.g. cardiomyopathy,  
 CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
 CC treating or preventing diseases such as inflammation, autoimmune  
 CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
 CC (AIDS), obesity, asthma, immunoglobulin (IgA) nephropathy, cirrhosis,

CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
 CC and epilepsy. Accordingly, these molecules have many activities including  
 CC cytostatic, cardiant, antidiabetic, antitumor, immunosuppressive, antiallergic,  
 CC haemostatic, anti-HIV, antidiabetic, antiatherosclerotic, anorectic,  
 CC antiaesthetic, nephrotoxic, antihypertensive, hepatotoxic, antiparasitic,  
 CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,  
 CC relaxant and antiviral. In addition, they are useful in screening  
 CC assays to identify small molecules that modulate or inhibit, for example,  
 CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
 CC used as in chromosome mapping, tissue typing, preventive medicine and  
 CC pharmacogenomics. This polypeptide is a homologue of a human NOXV protein  
 CC of the invention.  
 CC  
 XX  
 SO Sequence 855 AA;  
 Query Match 99.4%; Score 4653; DB 5; Length 855;  
 Best Local Similarity 99.5%; Pred. No. 2,4e-307;  
 Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MGSBRARKGGGGPRDFAGLKYNSRHKVNGLEGVEFLPVNNVKKYKGGPRMVLAA 60  
 DB 1 MGSBRARKGGGGPRDFAGLKYNSRHKVNGLEGVEFLPVNNVKKYKGGPRMVLAA 60  
 QY VILGLLVILGIGFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSSTEFVSLASKV 120  
 DB VILGLLVILGIGFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSSTEFVSLASKV 120  
 QY 121 KDALKLYSGVPLGIPYHKESAVTAPEGSVIAIYMSSEFIPQHLVEARVAEEVVM 180  
 DB 121 KDALKLYSGVPLGIPYHKESAVTAPEGSVIAIYMSSEFIPQHLVEARVAEEVVM 180  
 QY 121 LPPRARSLSFVTVSVAFPTDSKTQVORTODNSCSFGLHARGVLMFTTGGPDSRYPA 240  
 DB 181 LPPRARSLSFVTVSVAFPTDSKTQVORTODNSCSFGLHARGVLMFTTGGPDSRYPA 240  
 QY 241 HARCQWALRGDADSVLTPRSFDLASCDERSGLVTVNTLSPMERHALVOLCGTYPPS 300  
 DB 241 HARCQWALRGDADSVLTPRSFDLASCDERSGLVTVNTLSPMERHALVOLCGTYPPS 300  
 QY 301 YNLTFHSSQVNLITLITNTERRHPGFEATFPOLPRMSCGGRLLRKAQGTFFSPYPGHY 360  
 DB 301 YNLTFHSSQVNLITLITNTERRHPGFEATFPOLPRMSCGGRLLRKAQGTFFSPYPGHY 360  
 QY 361 PPNIDCTWNTIEVNNQNVKRFKPFYILBPGVAGTCPCQVYVINGKYGESGOFVNTS 420  
 DB 361 PPNIDCTWNTIEVNNQNVKRFKPFYILBPGVAGTCPCQVYVINGKYGESGOFVNTS 420  
 QY 421 NSNKITVRFHSDSYTDTGFLAAYLSYSSDPGQFTCRGRCIRKELACDGMADCTDH 480  
 DB 421 NSNKITVRFHSDSYTDTGFLAAYLSYSSDPGQFTCRGRCIRKELACDGMADCTDH 480  
 QY 481 SDELINSCDAAGHOFCKNFKCPFLFWCDVSNCGMNSDEQGSCEPQTRCSCNGKLSK 540  
 DB 481 SDELINSCDAAGHOFCKNFKCPFLFWCDVSNCGMNSDEQGSCEPQTRCSCNGKLSK 540  
 QY 541 SQQNGKDDGSDASCPKVVVCTKRTYCLNGSLCISKNPEDGKEDSDGSDSK 600  
 DB 541 SQQNGKDDGSDASCPKVVVCTKRTYCLNGSLCISKNPEDGKEDSDGSDSK 600  
 QY 601 DCDGGLSFTROARVVGTDADGEMPMQVSLHALGSHICGSLISPMNLVSAHCYID 660  
 DB 601 DCDGGLSFTROARVVGTDADGEMPMQVSLHALGSHICGSLISPMNLVSAHCYID 660  
 QY 661 DRGRFSDPQWTAFLGLHDQGRSAPGVOERLKRKIIISHPFNDFTFDYDILBLEKP 720  
 DB 661 DRGRFSDPQWTAFLGLHDQGRSAPGVOERLKRKIIISHPFNDFTFDYDILBLEKP 720  
 QY 721 AYSWWRPCLPDASHVPAKAIWVGSHQYSGTGLILQKEIRVINTTCENIL 780  
 DB 721 AYSWWRPCLPDASHVPAKAIWVGSHQYSGTGLILQKEIRVINTTCENIL 780  
 QY 781 PQQITPRMVCVGLSGVDSGQDSSGFLSVBADGRIFGAGVSWGDCAGBNKQGVYT 840  
 DB 781 PQQITPRMVCVGLSGVDSGQDSSGFLSVBADGRIFGAGVSWGDCAGBNKQGVYT 840

Db	781	PQQTTPMCMCVGLSGVDSGQDSGRLSSVEADGRIFQAGVSWGDCQARRKPGVTT	844
Oy	841	RLPLFRDWIKENTGV	855
Db	841	RLPLFRDWIKENTGV	855
RESULT 16			
AD116875			
ID	AD116875	standard; protein; 855 AA.	
XX	AC		
XX	AD116875;		
DT	15-APR-2004	(first entry)	
XX			
DE	Human NOXV protein homologue SegID 411.		
XX			
KW	human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;		
KW	inflammation; autoimmune disorder; allergy; blood disorder;		
KW	acquired immunodeficiency syndrome; AIDS; obesity; asthma;		
KW	immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;		
KW	Alzheimer's disease; infection; str.		
XX			
OS	Homo sapiens.		
XX			
PN	MO200268649-A2.		
XX			
PD	06-SEP-2002.		
XX			
PF	31-JAN-2002;	2002MO-US0002785.	
XX			
PR	31-JAN-2001;	2001US-0265395P.	
PR	31-JAN-2001;	2001US-0265412P.	
PR	31-JAN-2001;	2001US-0265514P.	
PR	31-JAN-2001;	2001US-0265517P.	
PR	02-FEB-2001;	2001US-0266406P.	
PR	05-FEB-2001;	2001US-0266767P.	
PR	07-FEB-2001;	2001US-0266975P.	
PR	07-FEB-2001;	2001US-0267057P.	
PR	08-FEB-2001;	2001US-0267459P.	
PR	09-FEB-2001;	2001US-0267823P.	
PR	15-FEB-2001;	2001US-0268974P.	
PR	26-FEB-2001;	2001US-0271664P.	
PR	27-FEB-2001;	2001US-0271839P.	
PR	27-FEB-2001;	2001US-0271855P.	
PR	02-MAR-2001;	2001US-0272788P.	
PR	02-MAR-2001;	2001US-0273046P.	
PR	14-MAR-2001;	2001US-0275925P.	
PR	14-MAR-2001;	2001US-0275947P.	
PR	14-MAR-2001;	2001US-0275950P.	
PR	14-MAR-2001;	2001US-0275989P.	
PR	15-MAR-2001;	2001US-0276448P.	
PR	15-MAR-2001;	2001US-0276450P.	
PR	16-MAR-2001;	2001US-0276397P.	
PR	16-MAR-2001;	2001US-0276768P.	
PR	20-MAR-2001;	2001US-0278652P.	
PR	26-MAR-2001;	2001US-0278775P.	
PR	26-MAR-2001;	2001US-0278778P.	
PR	29-MAR-2001;	2001US-0279882P.	
PR	29-MAR-2001;	2001US-0279884P.	
PR	30-MAR-2001;	2001US-0280147P.	
PR	11-APR-2001;	2001US-0282992P.	
PR	11-APR-2001;	2001US-0283083P.	
PR	20-APR-2001;	2001US-0285133P.	
PR	23-APR-2001;	2001US-0285749P.	
PR	03-MAY-2001;	2001US-0288337P.	
PR	03-MAY-2001;	2001US-0288504P.	
PR	29-MAY-2001;	2001US-02940647P.	
PR	30-MAY-2001;	2001US-0294473P.	
PR	08-JUN-2001;	2001US-0296964P.	
PR	18-JUN-2001;	2001US-0298959P.	
PR	19-JUN-2001;	2001US-0299334P.	
PR	13-AUG-2001;	2001US-0312020P.	

PR	16-AUG-2001;	2001US-0312889P.
XX	16-AUG-2001;	2001US-0312908P.
PR	21-AUG-2001;	2001US-0313390P.
PR	28-AUG-2001;	2001US-0315470P.
PR	31-AUG-2001;	2001US-0316447P.
PR	07-SEP-2001;	2001US-0318115P.
PR	07-SEP-2001;	2001US-0318118P.
PR	12-SEP-2001;	2001US-0318749P.
PR	18-SEP-2001;	2001US-0323379P.
PR	18-OCT-2001;	2001US-0330245P.
PR	18-OCT-2001;	2001US-0330308P.
PR	14-NOV-2001;	2001US-0332701P.
PA	(CURA-) CURAGEN CORP.	
XX	Tchernev VT, Spytek KA, Zernhusen BD, Paturajan M, Shinkets RA;	
PI	Li L, Gangoli EA, Padigar M, Anderson DW, Rastelli L, Miller CE;	
PI	Gutlach W, Tappier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CEA;	
PI	Furtak K, Grose WM, Alebrook JP, Lepley DW, Rieger DK, Burgess CE;	
XX	WPI; 2002-706998/76.	
DR	New NOXV polypeptides and nucleic acids, useful for preventing or	
PT	treating NOXV-associated disorders, e.g. cancer, cardiomyopathy,	
PT	atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or	
PT	pharmacogenomics.	
XX	Disclosure; SEQ ID NO 411; 1498bp; English.	
PS	This invention relates to a novel nucleic acid, and encoded polypeptides	
XX	thereof, which have properties related to the stimulation of biochemical	
CC	or physiological responses in a cell, tissue, organ or organism.	
CC	Specifically, it refers to the use of biologically active fragments for	
CC	diagnostic and prognostic assays and furthermore in the treatment of	
CC	diverse pathological conditions. The present invention describes novel	
CC	human and murine NOXV proteins, as well as methods to modulate their	
CC	expression using antisense oligos, ribozymes and peptide nucleic acids.	
CC	The NOXV polypeptides, polymucleotides and antibodies are useful in	
CC	treating or preventing NOXV-associated disorders, e.g., cardiomyopathy,	
CC	atherosclerosis, cancer and diabetes. Furthermore, they may be used in	
CC	treating or preventing diseases such as inflammation, autoimmune	
CC	disorders, allergies, blood disorders, acquired immunodeficiency syndrome	
CC	(AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,	
CC	arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy,	
CC	and epilepsy. Accordingly, these molecules have many activities including	
CC	cyclostatic, cardiant, antiinflammatory, immunosuppressive, antileptic,	
CC	haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,	
CC	antiasmatic, nephrotropic, antiarthritic, hepatotropic,	
CC	neuroprotective, neurotropic, antibacterial, vincinide, antiparasitic,	
CC	relaxant and anticoagulant. In addition, they are useful in screening	
CC	assays to identify small molecules that modulate or inhibit, for example,	
CC	neurgenesis, wound healing and angiogenesis. The nucleic acids are also	
CC	used as in chromosome mapping, tissue typing, preventive medicine and	
CC	pharmacogenomics. This polypeptide is a homologue of a human NOXV protein	
CC	of the invention.	
SQ	Sequence 855 AA:	
OY	Query Match	99.4%; Score 4653; DB 5; Length 855;
DB	Best Local Similarity	99.5%; Pred. No. 2.4e-307;
	Matches	851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY	1	MGSDPARKGGGGPPDFFAGLKYNSRHEENVNGLESGVEFLPVNNNYKKVKHGPGRRVVTLAA 60
DB	1	MGSDFARKGGGGPPDFFAGLKYNSRHEKVNLBEGVEFLPVNNNYKKVKHDPGRWRVTLLAA 60
OY	VLIGILLVLIGTIGLVNHLQRYADRVQGVKKVGKRIITNEENVDAEYENSSTEFVSLASKY 120	
DB	61	VLIIGILLVLIGTIGLVNHLQRYADRVQGVKKVGKRIITNEENVDAEYENSSTEFVSLASKY 120
OY	121	KDALKLTVSGVPFIQGYPHKESAVTAFFSEGSVIAYYWSEFSIPQHLVEBAERVMAEERYVM 180
DB	121	KDALKLTVSGVPFIQGYPHKESAVTAFFSEGSVIAYYWSEFSIPQHLVEBAERVMAEERYVM 180
DB	121	KDALKLTVSGVPFIQGYPHKESAVTAFFSEGSVIAYYWSEFSIPQHLVEBAERVMAEERYVM 180

QY 181 LPPRARSLSKSVTVSVVAFPTDSKTQVORTODNSCSFGIHAAGVELMRPTTGPDPSPYA 240  
 DB 181 LPPRARSLSKSVTVSVVAFPTDSKTQVORTODNSCSFGIHAAGVELMRPTTGPDPSPYA 240  
 QY 241 HARCQMLARGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300  
 DB 241 HARCQMLARGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300  
 QY 301 YNLTFFSSQNVLLITLITNTERRHPGFEATFPOLPRMSSCGRLRKAQGTFFNSPYYPGHY 360  
 DB 301 YNLTFFSSQNVLLITLITNTERRHPGFEATFPOLPRMSSCGRLRKAQGTFFNSPYYPGHY 360  
 QY 361 PPNIDCTWNIENVNNQHVKKRFFKFFYLLEBPGVAGTCPKDVEINGEKYCGERSQFVVT 420  
 DB 361 PPNIDCTWNIENVNNQHVKKRFFKFFYLLEBPGVAGTCPKDVEINGEKYCGERSQFVVT 420  
 QY 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSDPGCGFTCRGTGRCIRKELRCGDMADCTDH 480  
 DB 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSDPGCGFTCRGTGRCIRKELRCGDMADCTDH 480  
 QY 481 SDELINCSADAGHPTCKNKKFCKPLFWVCDVNDGSDNSDEQSCSPAQTFRCSNGKCLSK 540  
 DB 481 SDELINCSADAGHPTCKNKKFCKPLFWVCDVNDGSDNSDEQSCSPAQTFRCSNGKCLSK 540  
 QY 541 SQQNGKDDCGSDGSDASCPKNNVVTCTKHYRCLNGICLSKNPEDCKEDCGSDGDEK 600  
 DB 541 SQQNGKDDCGSDGSDASCPKNNVVTCTKHYRCLNGICLSKNPEDCKEDCGSDGDEK 600  
 QY 601 DCDGSLRFTFOARVVGTDDEGEWMPQVSLHALGQHIICGASLISPNMLVSAHACYID 660  
 DB 601 DCDGSLRFTFOARVVGTDDEGEWMPQVSLHALGQHIICGASLISPNMLVSAHACYID 660  
 QY 661 DRGRYSYDPTQWIAFLGLHDQSQRAPGVQRRRLKRIISHPFNDFTFYDIALLEBKP 720  
 DB 661 DRGRYSYDPTQWIAFLGLHDQSQRAPGVQRRRLKRIISHPFNDFTFYDIALLEBKP 720  
 QY 721 AEYSMSWRPITCLPASHVFPKGAIMWTGMHTQYGGGALLDKGEIRVINOTTENL 780  
 DB 721 AEYSMSWRPITCLPASHVFPKGAIMWTGMHTQYGGGALLDKGEIRVINOTTENL 780  
 QY 781 PQQITPRMVCVGFSGVDSQCGSDGGLSVLEADGRIFGAGVSWGDCAGRNKPGVYT 840  
 DB 781 PQQITPRMVCVGFSGVDSQCGSDGGLSVLEADGRIFGAGVSWGDCAGRNKPGVYT 840  
 QY 841 RLPLFRDMIKENTGV 855  
 DB 841 RLPLFRDMIKENTGV 855

RESULT 17  
 ABP56619  
 ID ABP56619 standard; protein; 855 AA.  
 XX  
 AC ABP56619;  
 XX  
 DT 24-MAR-2003 (first entry)  
 XX  
 DE Human membrane-type serine protease MTSP1 protein SEQ ID NO:2.  
 XX  
 KM Human: membrane-type serine protease; enzyme; MTSP10; cytochrome c;  
 KM type-II membrane-type serine protease; neoplastic disease; tumour; MTSP1;  
 KM matrixase.  
 XX  
 OS Homo sapiens.  
 OS  
 PN MO200292841-A2.  
 XX  
 PD 21-NOV-2002.  
 XX  
 PF 14-MAY-2002; 2002WO-US015332.  
 XX  
 PR 14-MAY-2001; 2001US-0291001P.

XX  
 PA (CORV-) CORVAS INT INC.  
 XX  
 PI Madison EL, Yen J;  
 XX  
 DR WPI; 2003-129309/12.  
 XX  
 DR N-PSDB; AB222450.  
 PT  
 PT New polypeptides comprising the protease domain of a type-II membrane-  
 PT type serine protease (MTSP10), or its mutants, useful for diagnosing  
 PT neoplasms or malignancies, or for screening for MTSP10 inhibitors for  
 PT treating such diseases.  
 XX  
 PS Disclosure; Page 181-183; 198pp; English.  
 CC  
 CC The present invention describes a polypeptide comprising a purified  
 CC single or two chain polypeptide, which comprises the protease domain of a  
 CC type-II membrane-type serine protease (MTSP10) or its catalytically  
 CC active portion, or a mutant of it, where up to 50 % of the amino acids  
 CC are replaced with another amino acid, and the resulting polypeptide is a  
 CC single chain or two chain polypeptide that has a catalytic activity of at  
 CC least 1-10 % of the unmutated polypeptide. MTSP10 has cytostatic  
 CC activity. The polypeptide containing the protease domain of the MTSP10 is  
 CC useful for detecting a neoplastic disease, and for diagnosing the  
 CC presence of a pre-malignant lesion, a malignancy, or other pathologic  
 CC condition in a subject, or monitoring tumour (e.g. breast, cervix,  
 CC prostate, lung, ovary or colon tumour) progression and/or therapeutic  
 CC effectiveness. An inhibitor of the polypeptide containing the protease  
 CC domain of MTSP10 is useful for treating or preventing neoplastic disease  
 CC in a mammal. An inhibitor of the activation or cleavage of the zymogen form  
 CC of the MTSP10 polypeptide is useful for inhibiting tumour initiation,  
 CC growth or progression, or treating (pre-)malignant conditions of the e.g.  
 CC breast, cervix, prostate, lung, ovary or colon. The present sequence  
 CC represents human MTSP1 (also known as matrixase), which is used in an  
 CC example from the present invention  
 XX  
 SQ Sequence 855 AA:  
 Query Match 99.4%; Score 4653; DB 6; Length 855;  
 Best Local Similarity 99.5%; Pred. No. 2,4e-307;  
 Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MGSDBARKGGGGPKDPCAGLAKTNSRHKVNGLEBGEVFLPVNNVKKYKHPGRMVLAA 60  
 DB 1 MGSDBARKGGGGPKDPCAGLAKTNSRHKVNGLEBGEVFLPVNNVKKYKHPGRMVLAA 60  
 QY 61 VILGLLVILGIGFLVWHLQYRDVAVQVKNQGYRITNENFVDAYENSSTEFVSLASKV 120  
 DB 61 VILGLLVILGIGFLVWHLQYRDVAVQVKNQGYRITNENFVDAYENSSTEFVSLASKV 120  
 QY 121 KDALKILYSGVPLFGPYHKESAVTAPSEGSVIAIYNSSEFIPQHLVBEARVVAEERVVM 180  
 DB 121 KDALKILYSGVPLFGPYHKESAVTAPSEGSVIAIYNSSEFIPQHLVBEARVVAEERVVM 180  
 QY 181 LPPRARSLSKSVTVSVVAFPTDSKTQVORTODNSCSFGIHAAGVELMRPTTGPDPSPYA 240  
 DB 181 LPPRARSLSKSVTVSVVAFPTDSKTQVORTODNSCSFGIHAAGVELMRPTTGPDPSPYA 240  
 QY 241 HARCQMLARGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300  
 DB 241 HARCQMLARGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300  
 QY 301 YNLTFFSSQNVLLITLITNTERRHPGFEATFPOLPRMSSCGRLRKAQGTFFNSPYYPGHY 360  
 DB 301 YNLTFFSSQNVLLITLITNTERRHPGFEATFPOLPRMSSCGRLRKAQGTFFNSPYYPGHY 360  
 QY 361 PPNIDCTWNIENVNNQHVKKRFFKFFYLLEBPGVAGTCPKDVEINGEKYCGERSQFVVT 420  
 DB 361 PPNIDCTWNIENVNNQHVKKRFFKFFYLLEBPGVAGTCPKDVEINGEKYCGERSQFVVT 420  
 QY 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSDPGCGFTCRGTGRCIRKELRCGDMADCTDH 480  
 DB 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSDPGCGFTCRGTGRCIRKELRCGDMADCTDH 480



QY 481 SDELINCSDAAGHQTCKNKECKPLFWVCDSYNDGSDSEOGSCSPAQTRCNSNGKCLSK 540  
DB 481 SDELINCSDAAGHQTCKNKECKPLFWVCDSYNDGSDSEOGSCSPAQTRCNSNGKCLSK 540  
QY 541 SQQCNKGKDDCGSDSDASCPKXNVVVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600  
DB 541 SQQCNKGKDDCGSDSDASCPKXNVVVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600  
QY 601 DCDGGLSFTROARVVGTDADGEMPMOYSLHALGQGHICGASLISPMNLVSAHCYID 660  
DB 601 DCDGGLSFTROARVVGTDADGEMPMOYSLHALGQGHICGASLISPMNLVSAHCYID 660  
QY 661 DRGFRYSPTQMTAFGLHDQSORSAPOVERRLKRIISHPFNDFTFYDIALLELEKP 720  
DB 661 DRGFRYSPTQMTAFGLHDQSORSAPOVERRLKRIISHPFNDFTFYDIALLELEKP 720  
QY 721 AEYSSMWPRPICLPDASHVFPAGKAIWYTGWHTOYGGTALLIÖKEIRIVINÖTTENLL 780  
DB 721 AEYSSMWPRPICLPDASHVFPAGKAIWYTGWHTOYGGTALLIÖKEIRIVINÖTTENLL 780  
QY 781 PQQITPRMNCVGLSGVDSCQDSDGGPLSSVEADGRIIFGAGVSWGDGACGRKPGVYT 840  
DB 781 PQQITPRMNCVGLSGVDSCQDSDGGPLSSVEADGRIIFGAGVSWGDGACGRKPGVYT 840  
QY 841 RLPLFRDWIKENTGV 855  
DB 841 RLPLFRDWIKENTGV 855

## RESULT 18

AAO30146 standard; protein; 855 AA.

AAO30146;

03-SEP-2003 (first entry)

Human membrane-type serine protease MTSP1 protein.

KW Serine protease 17; CYPSP1; tumour; cancer; anticense therapy; prostate;  
KW breast; cervix; lung; ovary; colon; gene therapy; human; enzyme; MTSP1;  
KW membrane-type serine protease; matrixplase.

Homo sapiens.

W02003044179-A2.

30-MAY-2003.

20-NOV-2002; 2002MO-US037626.

20-NOV-2001; 2001US-0332015P.

(CORV-) CORVAS INT INC.

Madison EL, Ong EO;

WPI; 2003-449816/42.

N-PSDB; AAL60792.

PT New substantially purified serine protease 17 polypeptide and encoding  
PT nucleic acid, useful for diagnosing and treating tumor conditions and/or  
PT cancer, particularly of the breast, cervix, prostate, lung, ovary or  
PT colon.

Disclousure; Page 175-177, 189pp, English.

CC The invention relates to serine protease 17 polypeptide designated CYPSP17  
CC and its corresponding nucleic acid sequence. The invention also relates  
CC to a method using CYPSP17 protein to identify compounds that modulate its  
CC protease activity. The method is useful for preventing, diagnosing and  
CC treating disorders related to the serine protease 17 activity, such as

CC tumour conditions and/or cancer, particularly of the breast, prostate,  
CC cervix, lung, ovary or colon. CYPSP17 DNA is used in gene therapy and in  
CC anticense therapy. The present sequence is human membrane-type serine  
CC protease MTSP1 (also called matrixplase) protein. This sequence is used to  
CC illustrate the method of the invention

SQ Sequence 855 AA;

Query Match 99.4%; Score 4653; DB 6; Length 855;

Best Local Similarity 99.5%; Pred. No. 2,4e-307;

Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDDARKGGGGKDFGAGLKYNSRHEKYNLEEGVEFLPVNNKKYKRGKRWVLLAA 60  
DB 1 MGSDDARKGGGGKDFGAGLKYNSRHEKYNLEEGVEFLPVNNKKYKRGKRWVLLAA 60  
QY 61 VILGLLVLLIGLGFVWHLQYRDVROKXNGYMRITNPNFVAYENSSTEFVLSASKY 120  
DB 61 VILGLLVLLIGLGFVWHLQYRDVROKXNGYMRITNPNFVAYENSSTEFVLSASKY 120  
QY 121 KDALKILYSGVPIGPIYHKEASAVTAFSEGSVLAAYSEFSIPQHLVEAEARVMAEERVYM 180  
DB 121 KDALKILYSGVPIGPIYHKEASAVTAFSEGSVLAAYSEFSIPQHLVEAEARVMAEERVYM 180  
QY 181 LPPRARSLSFVVTSVVAFPTDSKTYQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240  
DB 181 LPPRARSLSFVVTSVVAFPTDSKTYQRTQDNSCSFGLHARGVELMRFTTPGFPDSPYPA 240  
QY 241 HARCQWALRGDADSVSLTFRSPDLASCDERGSIDLTVYNTLSBMPHALLVOLCGYTPPS 300  
DB 241 HARCQWALRGDADSVSLTFRSPDLASCDERGSIDLTVYNTLSBMPHALLVOLCGYTPPS 300  
QY 301 YNLTFFSQNVLLITLITWTERHPGEATFFQLPRMSSCGRLRKAOGTFNSPYYPGHY 360  
DB 301 YNLTFFSQNVLLITLITWTERHPGEATFFQLPRMSSCGRLRKAOGTFNSPYYPGHY 360  
QY 361 PENIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPODYVINEBKTCGERSGQVVTIS 420  
DB 361 PENIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCPODYVINEBKTCGERSGQVVTIS 420  
QY 421 NSKITTIVRPHSDSYNDTGFLEVLSYDSDPPQGFORTGTGCIKKEIRCDQMACTDH 480  
DB 421 NSKITTIVRPHSDSYNDTGFLEVLSYDSDPPQGFORTGTGCIKKEIRCDQMACTDH 480  
QY 481 SDELINCSDAAGHQTCKNKECKPLFWVCDSYNDGSDSEOGSCSPAQTRCNSNGKCLSK 540  
DB 481 SDELINCSDAAGHQTCKNKECKPLFWVCDSYNDGSDSEOGSCSPAQTRCNSNGKCLSK 540  
QY 541 SQQCNKGKDDCGSDSDASCPKXNVVVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600  
DB 541 SQQCNKGKDDCGSDSDASCPKXNVVVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600  
QY 601 DCDGGLSFTROARVVGTDADGEMPMOYSLHALGQGHICGASLISPMNLVSAHCYID 660  
DB 601 DCDGGLSFTROARVVGTDADGEMPMOYSLHALGQGHICGASLISPMNLVSAHCYID 660  
QY 661 DRGFRYSPTQMTAFGLHDQSORSAPOVERRLKRIISHPFNDFTFYDIALLELEKP 720  
DB 661 DRGFRYSPTQMTAFGLHDQSORSAPOVERRLKRIISHPFNDFTFYDIALLELEKP 720  
QY 721 AEYSSMWPRPICLPDASHVFPAGKAIWYTGWHTOYGGTALLIÖKEIRIVINÖTTENLL 780  
DB 721 AEYSSMWPRPICLPDASHVFPAGKAIWYTGWHTOYGGTALLIÖKEIRIVINÖTTENLL 780  
QY 781 PQQITPRMNCVGLSGVDSCQDSDGGPLSSVEADGRIIFGAGVSWGDGACGRKPGVYT 840  
DB 781 PQQITPRMNCVGLSGVDSCQDSDGGPLSSVEADGRIIFGAGVSWGDGACGRKPGVYT 840  
QY 841 RLPLFRDWIKENTGV 855  
DB 841 RLPLFRDWIKENTGV 855

RESULT 19  
AAE29820  
ID AAE29820 standard; protein: 855 AA.  
XX  
AC AAE29820;  
XX  
DT 24-FEB-2003 (first entry)  
XX  
DE Human membrane-type serine protease 1 (MTSP1).  
XX  
KW Human; type II membrane-type serine protease 9; tumour; transgenic;  
KW type II transmembrane serine protease; enzyme; gene therapy; MTSP9;  
KW neoplastic disease; transgenic animal; membrane-type serine protease 1;  
KW TSP; MTSP1; matrilysin.  
XX  
OS Homo sapiens.  
XX  
PN W020277267-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-US009611.  
XX  
PR 27-MAR-2001; 2001US-0279228P.  
PR 15-MAY-2001; 2001US-0291501P.  
XX  
PA (CORV-) CORVAS INT INC.  
PI Madison BL, Ong EO;  
XX  
DR WPI: 2003-018940/01.  
DR N-PSDB; AAD47225.  
XX  
PT New substantially purified single or two-chain type II membrane-type  
PT serine protease 9 (MTSP9) polypeptide, useful for monitoring tumor  
PT progression, inhibiting tumor initiation, or treating a malignant or pre-  
PT malignant condition.  
XX  
PS Disclosure; Page 183-185; 199pp; English.  
XX  
CC The invention relates to type II membrane-type serine protease 9 (MTSP9)  
CC polypeptide and polymucloides. MTSP belongs to type II transmembrane  
CC serine protease (TSP) family. Sequences of the invention and their  
CC antibodies are useful for diagnosing, treating or preventing neoplastic  
CC disease in mammals. They are useful for monitoring tumor progression,  
CC inhibiting tumor initiation, growth or progression or treating malignant  
CC or pre-malignant conditions. Transgenic animals of the invention are  
CC useful in animal models of tumor initiation, growth and/or progression  
CC models. The invention is also useful in gene therapy. The present  
CC sequence is human membrane-type serine protease 1 (MTSP1). MTSP1 also  
CC referred as matrilysin is a member of the TSP family  
XX  
SQ Sequence 855 AA;

Query Match 99.4%; Score 4653; DB 6; Length 855;  
Best Local Similarity 99.5%; Pred. No. 2.4e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDRAKGGGPGDFAGLKYNSRHEKYNGLBGEVEFLPVNNVKKYKPGHVVVLA 60  
DB 1 MGSDRAKGGGPGDFAGLKYNSRHEKYNGLBGEVEFLPVNNVKKYKPGHVVVLA 60  
QY 61 VLIGLLLVLLGIGLVVHLYQYRVOKYKGYRINTENFVDAVENSNSTEYVSLASKV 120  
DB 61 VLIGLLLVLLGIGLVVHLYQYRVOKYKGYRINTENFVDAVENSNSTEYVSLASKV 120  
QY 121 KDALKLTVSGVPLGPIYKESAVTAFFSEGSVIATYMSSEFIPQHLVEBAEVMABERVVM 180  
DB 121 KDALKLTVSGVPLGPIYKESAVTAFFSEGSVIATYMSSEFIPQHLVEBAEVMABERVVM 180  
QY 181 LPPARSLKSFVTVSVVAFPTDSKTVQRTQDNCSCFGLHARGVELMRTTPGFPDSPYPA 240  
DB 181 LPPARSLKSFVTVSVVAFPTDSKTVQRTQDNCSCFGLHARGVELMRTTPGFPDSPYPA 240

QY 241 HARCQNALRGDADSVLSTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTTPS 300  
DB 241 HARCQNALRGDADSVLSTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTTPS 300  
QY 301 YNLTFFSSQNVLLITLTITNTERHPGFEATFQLPNMSCGGRLRKAQGTENSPYVGHY 360  
DB 301 YNLTFFSSQNVLLITLTITNTERHPGFEATFQLPNMSCGGRLRKAQGTENSPYVGHY 360  
QY 361 PNINDCTWNIIEVPNNQVKKRFFKFFYLLEPGVAGTCPKDYVEINGEKYCGESSQFYVTS 420  
DB 361 PNINDCTWNIIEVPNNQVKKRFFKFFYLLEPGVAGTCPKDYVEINGEKYCGESSQFYVTS 420  
QY 421 NSNKITVRFHSIDSYDTGTFLAELYSVSSDPGQGTCTGRCIRKELACDQADCTDH 480  
DB 421 NSNKITVRFHSIDSYDTGTFLAELYSVSSDPGQGTCTGRCIRKELACDQADCTDH 480  
QY 481 SDELNCSCDAGHOFCTCKNFKCKPLFWVCDSVNDGSDNDEQSCSPAQTFRCNSGKCLSK 540  
DB 481 SDELNCSCDAGHOFCTCKNFKCKPLFWVCDSVNDGSDNDEQSCSPAQTFRCNSGKCLSK 540  
QY 541 SQQCNKGDGCGSDDEASCPKRVNVVCTKATYRCLNGLCLSKNPNCEGKEDSDGSDK 600  
DB 541 SQQCNKGDGCGSDDEASCPKRVNVVCTKATYRCLNGLCLSKNPNCEGKEDSDGSDK 600  
QY 601 DCCGGLRSFTROARVVGCTADDEGEMPMOVSLLHALGGGHI CGASLISPMNLVSAHCTID 660  
DB 601 DCCGGLRSFTROARVVGCTADDEGEMPMOVSLLHALGGGHI CGASLISPMNLVSAHCTID 660  
QY 661 DRGFRYSDFQMTAFGLIHDQSORSAPGVQERLKRILSHPFNDFTFDYDIALLELEKP 720  
DB 661 DRGFRYSDFQMTAFGLIHDQSORSAPGVQERLKRILSHPFNDFTFDYDIALLELEKP 720  
QY 721 AEYSWVRPICLPDASHVFPAGKAIWVTGHTQYGGTGALLIQKGIKIRVINOCTCENL 780  
DB 721 AEYSWVRPICLPDASHVFPAGKAIWVTGHTQYGGTGALLIQKGIKIRVINOCTCENL 780  
QY 781 PQQITPRMNCVFLSGGVDSGCGSDGGLSSVADGRIFGAGVSWMGDCAGRNKPGVYT 840  
DB 781 PQQITPRMNCVFLSGGVDSGCGSDGGLSSVADGRIFGAGVSWMGDCAGRNKPGVYT 840  
QY 841 RLPLFRDMIKENTGV 855  
DB 841 RLPLFRDMIKENTGV 855

RESULT 20  
AAE29791  
ID AAE29791 standard; protein: 855 AA.  
XX  
AC AAE29791;  
XX  
DT 24-FEB-2003 (first entry)  
XX  
DE Human membrane-type serine protease, MTSP1.  
XX  
KW Human; serine protease 14; CYP14; cancer; malignancy; breast; colon;  
KW gene therapy; cytostatic; membrane-type serine protease; enzyme; MTSP1.  
XX  
OS Homo sapiens.  
XX  
PN W0200277263-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 20-MAR-2002; 2002WO-US009039.  
XX  
PR 22-MAR-2001; 2001US-0278166P.  
XX  
PA (CORV-) CORVAS INT INC.  
PI Madison BL, Yeh J;  
XX

DR WPI, 2003-018938/01.  
XX N-PSDB; AAD47180.  
PT New purified CVSP14 polypeptide and encoding nucleic acid molecule.  
PT Useful for diagnosis, preventing and/or treating disorders, such as  
PT cancers and malignancies of the breast, cervix, prostate, lung, ovary or  
PT colon.  
XX  
XX Disclosure; Page 171-173; 185pp; English.  
PS  
CC The invention relates to transmembrane serine protease 14 (CVSP14), its  
CC nucleic acid sequence and the method based on them. The methods and  
CC compositions of the invention are useful for diagnosing, preventing  
CC and/or treating conditions associated with the aberrant expression or  
CC activity of the CVSP14 polypeptide, such as cancers and malignancies of  
CC the breast, cervix, prostate, lung, ovary or colon. The methods are also  
CC useful for identifying compounds that will modulate the protease activity  
CC of CVSP14 polypeptide, and monitoring tumour progression and/or  
CC therapeutic effectiveness. CVSP14 DNA used in gene therapy. The present  
CC sequence is human membrane-type serine protease, MTSP1  
XX  
XX Sequence 855 AA:

Query Match 99.4%; Score 4653; DB 6; Length 855;  
Best Local Similarity 99.5%; Pred. No. 2,4e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MGSDRAKGGGGPGKDGAGLKYNSRHKVNGLEGEVFLPVNNVKYKXGPGRWVLLAA 60  
DB 1 MGSDRAKGGGGPGKDGAGLKYNSRHKVNGLEGEVFLPVNNVKYKXGPGRWVLLAA 60  
QY 61 VLIGLLVLLIGLIGLVVHLYQYRDVYQKXNGYMRITNENFVADYENSNSTEFVSLASKV 120  
DB 61 VLIGLLVLLIGLIGLVVHLYQYRDVYQKXNGYMRITNENFVADYENSNSTEFVSLASKV 120  
QY 121 KDALKLKYSGVPPFPGPHKSAVTAASEGSVIAAYWSEFSPQHLVEAEKRWAEERVVM 180  
DB 121 KDALKLKYSGVPPFPGPHKSAVTAASEGSVIAAYWSEFSPQHLVEAEKRWAEERVVM 180  
QY 181 LPPPARSLKSFVVTSSVVAFPDTSKYVORTODNSCFGILHARGVLMFTTGGFSDSPYPA 240  
DB 181 LPPPARSLKSFVVTSSVVAFPDTSKYVORTODNSCFGILHARGVLMFTTGGFSDSPYPA 240  
QY 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSDLVTYNTLSMEPHALVOLCGTYPSP 300  
DB 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSDLVTYNTLSMEPHALVOLCGTYPSP 300  
QY 301 YNLFHSSQNVLLITLTNTERRHPRGEATFPQLPRMSSCGRLRKAQGFNSPYFGHY 360  
DB 301 YNLFHSSQNVLLITLTNTERRHPRGEATFPQLPRMSSCGRLRKAQGFNSPYFGHY 360  
QY 361 PPNDICTWNIIEVPNNQHVAFKFFVYLEPGVPAQTCKDVEINERKCYGERSQFVVT 420  
DB 361 PPNDICTWNIIEVPNNQHVAFKFFVYLEPGVPAQTCKDVEINERKCYGERSQFVVT 420  
QY 421 NSNKITVRFHSDQSYDTDTGLAEVLSYDSDPCGQFTCRGRCIRKELRCGMDACTDH 480  
DB 421 NSNKITVRFHSDQSYDTDTGLAEVLSYDSDPCGQFTCRGRCIRKELRCGMDACTDH 480  
QY 481 SDELNCSGADAGHOTCKNKECKPLFWVCDSDVNDGSDSDGSCCPAQTFRCSNGKCLSK 540  
DB 481 SDELNCSGADAGHOTCKNKECKPLFWVCDSDVNDGSDSDGSCCPAQTFRCSNGKCLSK 540  
QY 541 SQQNGNGDGDGSDGDEASCPKVNVTCTKHTYRCLANGCLSKNGPECDGKEDGSDGDEK 600  
DB 541 SQQNGNGDGDGSDGDEASCPKVNVTCTKHTYRCLANGCLSKNGPECDGKEDGSDGDEK 600  
QY 601 DCDGGLRSFTRQARVVGCTADDEGEPMQVSLHALGQCHTIGASLISPMNLVSAHACYID 660  
DB 601 DCDGGLRSFTRQARVVGCTADDEGEPMQVSLHALGQCHTIGASLISPMNLVSAHACYID 660  
QY 661 DRGFRYSDPTQWTAFLGLHDQSQRSAQGVQERLRKRIISHPFNDPTFVDIALLELEKP 720  
DB 661 DRGFRYSDPTQWTAFLGLHDQSQRSAQGVQERLRKRIISHPFNDPTFVDIALLELEKP 720

DB 661 DRGFRYSDPTQWTAFLGLHDQSQRSAQGVQERLRKRIISHPFNDPTFVDIALLELEKP 720  
QY 721 AEYSSWVRPICLPDASHVPFAGKAIWWTGHTQYGGTGLILQKGEIRVYNQTCENLL 780  
DB 721 AEYSSWVRPICLPDASHVPFAGKAIWWTGHTQYGGTGLILQKGEIRVYNQTCENLL 780  
QY 781 PQQITPRMVCVGLSGVDSCQGDGSGPLSSVADGRIFGAGVVSNGDCAGNNKRGVYT 840  
DB 781 PQQITPRMVCVGLSGVDSCQGDGSGPLSSVADGRIFGAGVVSNGDCAGNNKRGVYT 840  
QY 841 RLPLFRDWIKENTGV 855  
DB 841 RLPLFRDWIKENTGV 855

RESULT 21  
ID ABP72376 standard; protein; 855 AA.  
XX  
AC ABP72376;  
XX  
DT 13-MAY-2003 (first entry)  
XX  
DE Transmembrane serine protease 1 (MTSP1).  
XX  
XX Human; transmembrane serine protease 1; MTSP1; MTSP20; enzyme;  
XX cytosolic; dermatological; cardiac; vulnary; ophthalmological;  
XX gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WC003004681-A2.  
XX  
PD 16-JAN-2003.  
XX  
PF 03-JUL-2002; 2002WC-US021208.  
XX  
PR 03-JUL-2001; 2001US-0302939P.  
XX  
PA (CORV-) CORVAS INT INC.  
XX  
PI Madison EL, Ong EO;  
XX  
XX WPI, 2003-239207/23.  
XX DR N-PSDB; ABZ58500.  
XX  
PT New type-II membrane-type serine protease 20 polypeptides, useful for  
PT preparing a membrane for diagnosing, treating or preventing cancer,  
PT dermatological disorders, aberrant wound repairs or crest syndromes.  
XX  
PS Disclosure; Page 198-200; 216pp; English.  
XX  
XX The present sequence is the protein sequence of human type II  
XX transmembrane serine protease 1 (MTSP1). The invention relates to novel  
XX human type II transmembrane serine protease 20 (MTSP20) (see ABP72374) and  
XX nucleic acids encoding it (see ABZ58499). Also claimed are methods of  
XX inhibiting tumour initiation, growth or progression by inhibiting MTSP20  
XX activity, and of treating or preventing a disease or disorder associated  
XX with undesired and/or uncontrolled angiogenesis or neovascularisation,  
XX especially undesired angiogenesis associated with solid neoplasms,  
XX vascular malformations and cardiovascular disorders, chronic inflammatory  
XX diseases, aberrant wound repairs, circulatory disorders, crest syndromes,  
XX dermatological disorders and ocular disorders using an inhibitor of MTSP20  
XX  
XX Sequence 855 AA:  
Query Match 99.4%; Score 4653; DB 6; Length 855;  
Best Local Similarity 99.5%; Pred. No. 2,4e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MGSDRAKGGGGPGKDGAGLKYNSRHKVNGLEGEVFLPVNNVKYKXGPGRWVLLAA 60  
DB 1 MGSDRAKGGGGPGKDGAGLKYNSRHKVNGLEGEVFLPVNNVKYKXGPGRWVLLAA 60

```

QY 61 VLIGLLVLIGIGLVHMLQYRDVYQKNGYMRITNENFVDAYENSNSTEFLASKV 120
DB 61 VLIGLLVLIGIGLVHMLQYRDVYQKNGYMRITNENFVDAYENSNSTEFLASKV 120
QY 121 KDALKLLYSQVPLGIPYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
DB 121 KDALKLLYSQVPLGIPYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
QY 181 LPPARSLKSFVVTSSVAAPFTDSKTVOQTQDNCSFGLHARGVLMFTTGPFPDSPYPA 240
DB 181 LPPARSLKSFVVTSSVAAPFTDSKTVOQTQDNCSFGLHARGVLMFTTGPFPDSPYPA 240
QY 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSMEPHALVOLCGTYPSPS 300
DB 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSMEPHALVOLCGTYPSPS 300
QY 301 YNLTFFHSSQNVLLITLTNTERRHHPGFATFPQLPRMSSCGGRLRKAQGTFSNPPYPGHY 360
DB 301 YNLTFFHSSQNVLLITLTNTERRHHPGFATFPQLPRMSSCGGRLRKAQGTFSNPPYPGHY 360
QY 361 PPNIIDCTWNIEVPNNQHVKRFKFFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVYTS 420
DB 361 PPNIIDCTWNIEVPNNQHVKRFKFFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVYTS 420
QY 421 NSNKITVRPHSDQSYDTDTGFLAEYLSYSDPCPGQFTCRTRGCIKRELRCDGMADCTDH 480
DB 421 NSNKITVRPHSDQSYDTDTGFLAEYLSYSDPCPGQFTCRTRGCIKRELRCDGMADCTDH 480
QY 481 SDELINCSGDAHQDTCKNKKFCCKPLFWVCDYVNDGSDNSDEQSCCPAQTFPRCSNGKCLSK 540
DB 481 SDELINCSGDAHQDTCKNKKFCCKPLFWVCDYVNDGSDNSDEQSCCPAQTFPRCSNGKCLSK 540
QY 541 SQQNGKXDDCGSDGDEASCPKRVNVTCTKTYRCLNGICLSKNPBCDCKEDCGSDGSEK 600
DB 541 SQQNGKXDDCGSDGDEASCPKRVNVTCTKTYRCLNGICLSKNPBCDCKEDCGSDGSEK 600
QY 601 DCDGGLSFTROAVVVGSTDADEGEWPMQVSLHALGQCHICGASLISPMNLVSAHACYID 660
DB 601 DCDGGLSFTROAVVVGSTDADEGEWPMQVSLHALGQCHICGASLISPMNLVSAHACYID 660
QY 661 DRGRYSDPTQMTAFGLIHDSORSAPVQERRLKRIISHPPFNDTFPDYDIALLELEKP 720
DB 661 DRGRYSDPTQMTAFGLIHDSORSAPVQERRLKRIISHPPFNDTFPDYDIALLELEKP 720
QY 721 AEYSMWVRPCLPDAHVFPAGKAIWWTGHTQYGGTGAIILOKGEIRVINQTTCEML 780
DB 721 AEYSMWVRPCLPDAHVFPAGKAIWWTGHTQYGGTGAIILOKGEIRVINQTTCEML 780
QY 781 PQOITPRMMCVGFLSGVDSQSGSGGSLSSVEADGRIFGAGVYVSWGCGCGARKKPGVYT 840
DB 781 PQOITPRMMCVGFLSGVDSQSGSGGSLSSVEADGRIFGAGVYVSWGCGCGARKKPGVYT 840
QY 841 RLPLFRDMIKENTGV 855
DB 841 RLPLFRDMIKENTGV 855

```

RESULT 22  
ADB97551  
ID ADB97551 standard; protein; 855 AA.

AC ADB97551;  
DT 04-DEC-2003 (first entry)  
DE Human MTSpl, SEQ ID NO:2.  
XX Human, type I transmembrane serine protease 1; MTSpl, matrix; serine  
XX protease; cleavage activation; diagnostic marker;  
KW neoplastic disorder; cancer; breast; cervix; prostate; lung; ovary;  
KW cytosolic; gene therapy; drug screening; tumour progression; monitoring;  
KW enzyme.

```

XX XX Homo sapiens.
OS OS
PN WO2003031585-A2.
XX 17-APR-2003.
XX 08-OCT-2002; 2002MO-US032417.
XX 09-OCT-2001; 2001US-0328530P.
XX (CORV-) CORVAS INT INC.
XX Madison EL, Yeh J;
XX WPI: 2003-393442/37.
XX N-PSDB; ADB97550.
XX New purified single- or two-chain polypeptide, useful for diagnosing,
XX preventing or treating cancer (e.g. colon cancer), comprises a protease
XX domain of a type-1 membrane-type serine protease 25 or its catalytically
XX active portion.
XX Disclosure; Page 179-181; 97pp; English.
XX The invention relates to human type I transmembrane serine protease 25
XX (MTSP25; ADB97555) and polypeptides derived from it (e.g., ADB97555). The
XX MTSP25 gene is located on chromosome 12. MTSP25 is a serine protease
XX which is activated by protease cleavage to yield a two-chain protease
XX comprising an A chain and a B chain linked by a disulphide bond. MTSP25
XX is expressed or is active in tumour cells, and can therefore be used as a
XX diagnostic marker for certain cancers. The invention also encompasses
XX nucleic acids encoding an MTSP25 polypeptide (ADB97564, ADB97571,
XX ADB97554); nucleic acid vectors and host cells comprising an MTSP25
XX polynucleotide; a MTSP25 knockout animal; and an antibody specific for
XX either the single chain (zymogen) or two-chain (activated) form of
XX MTSP25. MTSP25 polypeptides are useful in diagnosing, preventing or
XX treating neoplastic diseases, such as cancer of the breast, cervix,
XX prostate, lung, ovary or colon. MTSP25 polypeptides may also be used for
XX identifying compounds that modulate the protease activity of the
XX polypeptide and for monitoring tumour progression and/or therapeutic
XX effectiveness. The present sequence represents the related protein, MTSpl
XX (also known as matrilysin).
SQ Sequence 855 AA;
Query Match 99.4%; Score 4653; DB 7; Length 855;
Best Local Similarity 99.5%; Pred. No. 2,4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MGSDBARKGGGKPPDPAAGLKYNSRHEKNGLEGEVFLPVNNYKYEKGPGRWVLLAA 60
DB 1 MGSDBARKGGGKPPDPAAGLKYNSRHEKNGLEGEVFLPVNNYKYEKGPGRWVLLAA 60
QY 61 VLIGLLVLIGIGLVHMLQYRDVYQKNGYMRITNENFVDAYENSNSTEFLASKV 120
DB 61 VLIGLLVLIGIGLVHMLQYRDVYQKNGYMRITNENFVDAYENSNSTEFLASKV 120
QY 121 KDALKLLYSQVPLGIPYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
DB 121 KDALKLLYSQVPLGIPYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
QY 181 LPPARSLKSFVVTSSVAAPFTDSKTVOQTQDNCSFGLHARGVLMFTTGPFPDSPYPA 240
DB 181 LPPARSLKSFVVTSSVAAPFTDSKTVOQTQDNCSFGLHARGVLMFTTGPFPDSPYPA 240
QY 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSMEPHALVOLCGTYPSPS 300
DB 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSMEPHALVOLCGTYPSPS 300
QY 301 YNLTFFHSSQNVLLITLTNTERRHHPGFATFPQLPRMSSCGGRLRKAQGTFSNPPYPGHY 360
DB 301 YNLTFFHSSQNVLLITLTNTERRHHPGFATFPQLPRMSSCGGRLRKAQGTFSNPPYPGHY 360

```

QY 361 PPNIDCTWNIETVNNQHVKVRFEFFYLLEPVGAGTCPKDYVEINGEKYCGERSQFVVT  
PT 361 PPNIDCTWNIETVNNQHVKVSFKFFYLLEPVGAGTCPKDYVEINGEKYCGERSQFVVT  
Db 420  
QY 421 NSNKITVRFHSDOSYDTDTGFLAEYLSYDSDPGQFTCTGRCIRKELRCDGMADCTDH  
PT 421 NSNKITVRFHSDOSYDTDTGFLAEYLSYDSDPGQFTCTGRCIRKELRCDGMADCTDH  
Db 480  
QY 481 SDELNCSDAAGHQTCKNKKCKPLFWYCDVNDGSDNDEQSCSPAQTFRCSNGKCLSK  
PT 481 SDELNCSDAAGHQTCKNKKCKPLFWYCDVNDGSDNDEQSCSPAQTFRCSNGKCLSK  
Db 540  
QY 541 SQQCGKDDCGDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK  
PT 541 SQQCGKDDCGDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK  
Db 600  
QY 601 DCCGGLRSFTRQARVVGTDADGEGMPQVSLHALGQGHICGASLISPMVLVSAACCYID  
PT 601 DCCGGLRSFTRQARVVGTDADGEGMPQVSLHALGQGHICGASLISPMVLVSAACCYID  
Db 660  
QY 661 DRGFRYSDEPTQMTAFGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDIALLELKP  
PT 661 DRGFRYSDEPTQMTAFGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDIALLELKP  
Db 720  
QY 721 AEVSSWVRPILCPDASHVFPAGKAIWVTGHTQYGGTGALLIQKGEIRVINQTTCEML  
PT 721 AEVSSWVRPILCPDASHVFPAGKAIWVTGHTQYGGTGALLIQKGEIRVINQTTCEML  
Db 780  
QY 781 PQQITPRMVCVGLSGVDSCQDGGSPGLSSVEADGRIFGAGVSVWGDGCAQRKPGVYT  
PT 781 PQQITPRMVCVGLSGVDSCQDGGSPGLSSVEADGRIFGAGVSVWGDGCAQRKPGVYT  
Db 840  
QY 841 RLPLFRDMIKENTGV 855  
PT 841 RLPLFRDMIKENTGV 855  
Db 855 RLPLFRDMIKENTGV 855

RESULT 23  
AD10371  
ID AD10371 standard; protein; 855 AA.  
XX  
AC AD10371;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Human cell surface protease #1.  
XX  
KW therapeutic agent; plasmin; protease specific antigen; PSA;  
KW cell-surface protease-associated disease; cancer; ocular disease;  
KW cardiovascular disease; chronic inflammatory disease; wound;  
KW circulatory disorder; dermatological disorder; rheumatoid arthritis;  
KW psoriasis; diabetic retinopathy; pterygium;  
KW excimer laser surgery scarring; glaucoma filtering surgery scarring;  
KW macular degeneration; cress syndrome; solid neoplasia; vascular tumour;  
KW melanoma; Kaposi's sarcoma; human; cell surface protease.  
XX  
OS Homo sapiens.  
XX  
PN WO200295007-A2.  
XX  
PD 28-NOV-2002.  
XX  
PF 23-MAY-2002; 2002MO-US016819.  
XX  
PR 23-MAY-2001; 2001US-0293267P.  
XX  
PA (CORV-) CORVAS INT INC.  
XX  
PI Madison EL, Semple JE, Vlaauk GP, Kemp SJ, Komandla M, Stev DV;  
XX WPI; 2003-221280/21.  
DR N-PSDB; AD10370.

XX  
PT Novel conjugate useful for treating cell-surface protease-associated  
PT disease, comprises a therapeutic agent and a peptidic or nucleic acid  
PT substrate linked to it optionally by a peptidic linker.  
XX  
PS Claim 9, SEQ ID NO 2, 581bp, English.  
XX  
CC The invention comprises a conjugate that consists of a therapeutic agent  
CC and a peptide substrate (optionally linked via linker). The peptide  
CC substrate is proteolytically cleaved by a cell surface protease or a  
CC soluble, released or shed form of it, to liberate the therapeutic agent,  
CC the conjugate of the invention is not substantially cleaved by plasmin or  
CC protease specific antigen (PSA). The conjugate of the invention is useful  
CC for treating a cell-surface protease-associated disease such as: cancer,  
CC ocular diseases, cardiovascular diseases, chronic inflammatory diseases,  
CC wounds, circulatory disorders, dermatological disorders, rheumatoid  
CC arthritis, psoriasis, diabetic retinopathies, recurrence of pterygium,  
CC scarring from excimer laser surgery, scarring from glaucoma filtering  
CC surgery, macular degeneration, cress syndromes, solid neoplasms, vascular  
CC tumours, melanoma and Kaposi's sarcoma. The present amino acid sequence  
CC represents a human cell surface protease.  
XX  
SQ Sequence 855 AA:  
Query Match 99.4%; Score 4653; DB 7; Length 855;  
Best local Similarity 99.5%; Pred. No. 2, 4e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MSDDRARKGGGPKDAGAGIKVNSRHEKVGLEEGVEPLPVNNVKKYKGPGRWVTLA 60  
Db 1 MSDDRARKGGGPKDAGAGIKVNSRHEKVGLEEGVEPLPVNNVKKYKGPGRWVTLA 60  
QY 61 VLLGLLVLLGLIGFLVWHLQYRDVRYQKVGNGYMRITNENFVDAYENSTEFVSLASKV 120  
Db 61 VLLGLLVLLGLIGFLVWHLQYRDVRYQKVGNGYMRITNENFVDAYENSTEFVSLASKV 120  
QY 121 KDALKLLYSGVPLPGPHKESAVTAPSEGSVIAYYNSERSIPQHLVBEARVAEERVVM 180  
Db 121 KDALKLLYSGVPLPGPHKESAVTAPSEGSVIAYYNSERSIPQHLVBEARVAEERVVM 180  
QY 181 LPPRARSLSFVYTVSVAPPTDSKTQVORODNSCSGFLHARGVELMRFPTTGGPDSPPYA 240  
Db 181 LPPRARSLSFVYTVSVAPPTDSKTQVORODNSCSGFLHARGVELMRFPTTGGPDSPPYA 240  
QY 241 HARCQWALRGDADSVSLTFRSPDLASCDEKSGDLVTYVNTLSPMEPHALVOLCGTYPPS 300  
Db 241 HARCQWALRGDADSVSLTFRSPDLASCDEKSGDLVTYVNTLSPMEPHALVOLCGTYPPS 300  
QY 301 YNLTFFHSQNVLLITLITNTERHPGEATFPOLPMNSCGGRUKRAQGTFSPPYRGHY 360  
Db 301 YNLTFFHSQNVLLITLITNTERHPGEATFPOLPMNSCGGRUKRAQGTFSPPYRGHY 360  
QY 361 PPNIDCTWNIETVNNQHVKVRFEFFYLLEPVGAGTCPKDYVEINGEKYCGERSQFVVT 420  
Db 361 PPNIDCTWNIETVNNQHVKVSFKFFYLLEPVGAGTCPKDYVEINGEKYCGERSQFVVT 420  
QY 421 NSNKITVRFHSDOSYDTDTGFLAEYLSYDSDPGQFTCTGRCIRKELRCDGMADCTDH 480  
Db 421 NSNKITVRFHSDOSYDTDTGFLAEYLSYDSDPGQFTCTGRCIRKELRCDGMADCTDH 480  
QY 481 SDELNCSDAAGHQTCKNKKCKPLFWYCDVNDGSDNDEQSCSPAQTFRCSNGKCLSK 540  
Db 481 SDELNCSDAAGHQTCKNKKCKPLFWYCDVNDGSDNDEQSCSPAQTFRCSNGKCLSK 540  
QY 541 SQQCGKDDCGDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600  
Db 541 SQQCGKDDCGDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600  
QY 601 DCCGGLRSFTRQARVVGTDADGEGMPQVSLHALGQGHICGASLISPMVLVSAACCYID 660  
Db 601 DCCGGLRSFTRQARVVGTDADGEGMPQVSLHALGQGHICGASLISPMVLVSAACCYID 660  
QY 661 DRGFRYSDEPTQMTAFGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDIALLELKP 720

Db 661 DRGRYSDPTQWTAFLGLHDQSASAPGVDERRLKRIISHPFDFPFDDIALLELEK 720  
Qy 721 AEYSWVRPICLPASHVFPAGKAIWTTGHTGYGGTGAIILOKGEIRVINTTCENLL 780  
Db 721 AEYSWVRPICLPASHVFPAGKAIWTTGHTGYGGTGAIILOKGEIRVINTTCENLL 780  
Qy 781 PQQITPRMVCVGLSGVDSCQSGSGLSSVEADGRIFGAGVVSWDGCGAKRKPQVYT 840  
Db 781 PQQITPRMVCVGLSGVDSCQSGSGLSSVEADGRIFGAGVVSWDGCGAKRKPQVYT 840  
Qy 841 RLPLFRDMIKENTGV 855  
Db 841 RLPLFRDMIKENTGV 855  
RESULT 24  
ADG65326  
ID ADG65326 standard; protein; 855 AA.  
AC ADG65326;  
XX 11-MAR-2004 (first entry)  
XX Human MTSPI.  
XX human, type II membrane-type serine protease 12; MTSPI2; chromosome 19;  
KM protease domain; MTSPI2-P1; MTSPI2-P2; MTSPI2-P3; lung; oesophageal;  
KM prostate; colon; ovary; cervix; breast; pancreas; cancer; tumour;  
XX neoplastic condition.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH 615..855  
FT Domain /label = Protease domain  
XX  
PN W02003104394-A2.  
PD 18-DEC-2003.  
XX  
PF 21-MAY-2003; 2003WO-US016181.  
XX  
PR 21-MAY-2002; 2002US-0382851P.  
XX  
PA (DEND-) DENDREON SAN DIEGO LLC.  
XX  
PI Madison EL, Ong EO;  
XX  
DR WPI; 2004-062325/06.  
XX  
PT New type II membrane-type serine protease 12 proteins and nucleic acids,  
PT useful in diagnostics, particularly for diagnosing lung, esophageal,  
XX prostate, colon, ovary, cervix, breast and pancreas cancers.  
XX  
PS Disclosure; SEQ ID NO 2; 207bp; English.  
XX  
XX This sequence represents a human type II membrane-type serine protease 1  
CC (MTSPI). The MTSPI protein sequence was used in the isolation of the  
CC coding sequence of MTSPI2, which has been isolated to chromosome 19.  
CC MTSPI2 does not include the sequence of amino acids Arg-lys-His-Ileu-Pro-  
CC Arg-Pro-Ala ADG65347. The MTSPI2 coding sequence was identified by using  
CC the protein sequence of the protease domain of MTSPI to search the human  
CC HTGS database which produced three serine proteases, MTSPI2-P1, MTSPI2-  
CC P2 and MTSPI2-P3. EST's were identified which matched portions of  
CC MTSPI2-P1, MTSPI2-P2 and MTSPI2-P3. MTSPI2 polypeptides, proteins and  
CC nucleic acids are useful in diagnostics, particularly for diagnosing  
CC lung, oesophageal, prostate, colon, ovary, cervix, breast and pancreas  
CC cancers. These are useful in immunoassays to detect, prognose, diagnose,  
CC or monitor various conditions, diseases, and disorders affecting MTSPI2  
CC polypeptide expression, or monitor their treatment. Modulators of MTSPI2  
CC are useful for treating cancer, tumour and other neoplastic conditions.  
XX

Seq Sequence 855 AA;  
Query Match 99.4%; Score 4653; DB 8; Length 855;  
Best Local Similarity 99.5%; Pred. No. 2,4e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 MGSDBARKGGGGPDPFAGLKYNSRHEKVNGLSEGVFLPVNNVKYKKGKPGRWVTLAA 60  
Db 1 MGSDBARKGGGGPDPFAGLKYNSRHEKVNGLSEGVFLPVNNVKYKKGKPGRWVTLAA 60  
Qy 61 VLIGLLVLLIGLIGFLVWHLQYRDVRVQKXNGYRITNENFVAYENSNSTEFVSLASKV 120  
Db 61 VLIGLLVLLIGLIGFLVWHLQYRDVRVQKXNGYRITNENFVAYENSNSTEFVSLASKV 120  
Qy 121 KDALKLYSGVPLGPRHYKESAATAPSEGSVIAIYWESEFIPQHLVEARVMAEEVVM 180  
Db 121 KDALKLYSGVPLGPRHYKESAATAPSEGSVIAIYWESEFIPQHLVEARVMAEEVVM 180  
Qy 181 LPPRARSLSKFFVTVSVVAPPTDSKTVQRTODNSCSFGLHARGVELMRFITPGPPSPYPA 240  
Db 181 LPPRARSLSKFFVTVSVVAPPTDSKTVQRTODNSCSFGLHARGVELMRFITPGPPSPYPA 240  
Qy 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSPEMHALVOLCGTYPPS 300  
Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSPEMHALVOLCGTYPPS 300  
Qy 301 YNLTFFSSQVNLITLITNTERRHPRGEATFFQLPKNSCGGRLRKAQCTFNSPYRGGY 360  
Db 301 YNLTFFSSQVNLITLITNTERRHPRGEATFFQLPKNSCGGRLRKAQCTFNSPYRGGY 360  
Qy 361 PPNIDCTWNIIEVPNNQHVKRFKFFYLLBPGVPAGTCCPKDVEINGEKYGEERSQFVVS 420  
Db 361 PPNIDCTWNIIEVPNNQHVKRFKFFYLLBPGVPAGTCCPKDVEINGEKYGEERSQFVVS 420  
Qy 421 NSNKITVRFHSDSYTDYTGFLAEYLSYSSDPCPGQFTCRGRCIRKELRCDGADCTDH 480  
Db 421 NSNKITVRFHSDSYTDYTGFLAEYLSYSSDPCPGQFTCRGRCIRKELRCDGADCTDH 480  
Qy 481 SDELNCSGDAHQPTCKNFKCKPLFWVCDNVNDCGNSDDGSCCPAOTFRCSNGKLSK 540  
Db 481 SDELNCSGDAHQPTCKNFKCKPLFWVCDNVNDCGNSDDGSCCPAOTFRCSNGKLSK 540  
Qy 541 SQQNGKDDCGDSDDEASCPKNNVVTCTKATYRCLNGLCLSKNPECDGKEDCDSDGDEK 600  
Db 541 SQQNGKDDCGDSDDEASCPKNNVVTCTKATYRCLNGLCLSKNPECDGKEDCDSDGDEK 600  
Qy 601 DCDGRLSPTRQARVVGTDADGEMPMQVSLHALGGGHI CGASLISPNMLVSAAHCYID 660  
Db 601 DCDGRLSPTRQARVVGTDADGEMPMQVSLHALGGGHI CGASLISPNMLVSAAHCYID 660  
Qy 661 DRGRYSDPTQWTAFLGLHDQSASAPGVDERRLKRIISHPFDFPFDDIALLELEK 720  
Db 661 DRGRYSDPTQWTAFLGLHDQSASAPGVDERRLKRIISHPFDFPFDDIALLELEK 720  
Qy 721 AEYSWVRPICLPASHVFPAGKAIWTTGHTGYGGTGAIILOKGEIRVINTTCENLL 780  
Db 721 AEYSWVRPICLPASHVFPAGKAIWTTGHTGYGGTGAIILOKGEIRVINTTCENLL 780  
Qy 781 PQQITPRMVCVGLSGVDSCQSGSGLSSVEADGRIFGAGVVSWDGCGAKRKPQVYT 840  
Db 781 PQQITPRMVCVGLSGVDSCQSGSGLSSVEADGRIFGAGVVSWDGCGAKRKPQVYT 840  
Qy 841 RLPLFRDMIKENTGV 855  
Db 841 RLPLFRDMIKENTGV 855  
RESULT 25  
AD128861  
ID AD128861 standard; protein; 855 AA.  
AC AD128861;  
XX

DT 22-APR-2004 (first entry)  
XX Human matrixinase ( MTSP1) serine protease.  
XX Human, matrixinase; MTSP1; CVPSP16; cytostatic; gene therapy; tumour;  
KM marker; serine protease.  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT Domain 615..855  
/note="Protease domain"  
PN MO2004005471-A2.  
PD 15-JAN-2004.  
XX  
XX 01-JUL-2003; 2003MO-US020959.  
XX  
XX 02-JUL-2002; 2002US-0394347P.  
XX  
XX (DEND-) DENDREON SAN DIEGO LLC.  
PI Madison EL, Ong EO, Yeh J;  
XX  
XX MPI; 2004-099379/10.  
DR N-PSDB; ADI28860.  
XX  
XX  
XX New serine protease 16 polypeptides, useful for diagnosing, preventing or  
PT treating cancer (e.g. breast cancer) or for identifying compounds that  
PT may be used for modulating the polypeptides.  
XX  
XX Disclosure; SEQ ID NO 2; 205pp; English.  
XX  
XX The present sequence is the protein sequence of human MTSP1 (matrixinase),  
CC a serine protease that is expressed in epithelial cancer and normal  
CC tissue. In an example from the invention, the protein sequence of the  
CC protease domain of MTSP1 was used to search a human genome database. A  
CC protease was identified and designated serine protease 16 or CVPSP16.  
CC CVPSP16 cDNA ADI28863 was subsequently cloned from a human liver cDNA  
CC library by PCR. CVPSP16 polypeptides exhibit protease activity as a single  
CC chain or as a multi-chain form. Methods are provided for identifying  
CC compounds that modulate the protease activity. CVPSP16 polypeptides also  
CC serve as tumour markers.  
XX  
XX Sequence 855 AA:  
SQ  
Query Match 99.4%; Score 4653; DB 8; Length 855;  
Best Local Similarity 99.5%; Pred. No. 2.4e-307;  
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MGSDRAKGGGPGKDFAGLKYNSRHKVNGLBEGVFLPVNNVKKYKQGRWVLLA 60  
DB 1 MGSDRAKGGGPGKDFAGLKYNSRHKVNGLBEGVFLPVNNVKKYKQGRWVLLA 60  
QY 61 VLIGLLVLLIGLGLVHMLGYRDVRYQKXNNGYRITNENFVADYENSNSTEFVSLASKV 120  
DB 61 VLIGLLVLLIGLGLVHMLGYRDVRYQKXNNGYRITNENFVADYENSNSTEFVSLASKV 120  
QY 121 KDALKLLYSGVPLGPRHKSAYTAESEGSVIAATYNEEFSIPQULVEAEARVMAEERVVM 180  
DB 121 KDALKLLYSGVPLGPRHKSAYTAESEGSVIAATYNEEFSIPQULVEAEARVMAEERVVM 180  
QY 121 KDALKLLYSGVPLGPRHKSAYTAESEGSVIAATYNEEFSIPQULVEAEARVMAEERVVM 180  
DB 121 KDALKLLYSGVPLGPRHKSAYTAESEGSVIAATYNEEFSIPQULVEAEARVMAEERVVM 180  
QY 181 LPPRARSLSKFFVTVTSVAAPFDSKTVOPTODNSCSFGLHARGVLMFFTPGFPDSEYPA 240  
DB 181 LPPRARSLSKFFVTVTSVAAPFDSKTVOPTODNSCSFGLHARGVLMFFTPGFPDSEYPA 240  
QY 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSDLVTVTNLTSPMEBHALVOLCGTYPSP 300  
DB 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSDLVTVTNLTSPMEBHALVOLCGTYPSP 300  
QY 301 YNLTFFSSQWVLLITLTNTERRRPGFEATTFQLPKMSCCGRLRKAQGTFSNPTYYGHY 360  
DB 301 YNLTFFSSQWVLLITLTNTERRRPGFEATTFQLPKMSCCGRLRKAQGTFSNPTYYGHY 360

DB 301 YNLTFFSSQWVLLITLTNTERRRPGFEATTFQLPKMSCCGRLRKAQGTFSNPTYYGHY 360  
QY 361 PNIDCTWNIENVNNOHVKVRFPKFFYLLEBGVPACTCPKDYVEINSEKCYGERSQFVTS 420  
DB 361 PNIDCTWNIENVNNOHVKVRFPKFFYLLEBGVPACTCPKDYVEINSEKCYGERSQFVTS 420  
QY 421 NSNKITVRPHSDQSYTDTGFLAETLSYDSDPCPGQFTCTGACIRKELACDGMADCTDH 480  
DB 421 NSNKITVRPHSDQSYTDTGFLAETLSYDSDPCPGQFTCTGACIRKELACDGMADCTDH 480  
QY 481 SDELNCSCAAGHOFCTCKNFKCKPLFWVCDSVNDGNSDQGSCEPAQFRGSCNGKCLSK 540  
DB 481 SDELNCSCAAGHOFCTCKNFKCKPLFWVCDSVNDGNSDQGSCEPAQFRGSCNGKCLSK 540  
QY 541 SQQCNKGKDCGSDSDASCPKXNVVCTKHTYRCLNGLCLSKGNPCDQKEDCSDSDEK 600  
DB 541 SQQCNKGKDCGSDSDASCPKXNVVCTKHTYRCLNGLCLSKGNPCDQKEDCSDSDEK 600  
QY 601 DCDGGLRSFTQARVVVGTDADGEMPWQVSLHALQGHICGASLISPNVLVSAHACYTD 660  
DB 601 DCDGGLRSFTQARVVVGTDADGEMPWQVSLHALQGHICGASLISPNVLVSAHACYTD 660  
QY 661 DRGFRYSDPTQWTAFLGLHDQSORSA PGVERLKRILISHPFNDFTFYDIALLELEXP 720  
DB 661 DRGFRYSDPTQWTAFLGLHDQSORSA PGVERLKRILISHPFNDFTFYDIALLELEXP 720  
QY 721 AEYSWVRPICLPDASHVPAGKAIWVTGHTQYGGTALILQKEIRVINOCTENLL 780  
DB 721 AEYSWVRPICLPDASHVPAGKAIWVTGHTQYGGTALILQKEIRVINOCTENLL 780  
QY 781 PQOITPRMNCVGLSGVDSCQDSSGPISSVADGRIRAGVIVSGDGCAGNKKGYVT 840  
DB 781 PQOITPRMNCVGLSGVDSCQDSSGPISSVADGRIRAGVIVSGDGCAGNKKGYVT 840  
QY 841 RLPLFRDWIKENTGV 855  
DB 841 RLPLFRDWIKENTGV 855  
RESULT 26  
ADJ46895  
ID ADJ46895 standard; protein; 855 AA.  
XX  
AC ADJ46895;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
XX Human transmembrane serine protease (MTSP) polypeptide #1.  
XX  
XX Human; transmembrane serine protease; MTSP; cell surface protease;  
KM plasmin; prostate specific antigen; PSA; proliferative disease;  
KM cell-surface protease-associated disease; autoimmune disease;  
KM inflammatory disease; infectious disease; endocrine disease; cancer;  
KM ocular disorder; cardiovascular disorder; chronic inflammatory disease;  
KM wound; circulatory disorder; dermatological disorder; testecrosis;  
KM rheumatoid arthritis; psoriasis; diabetic retinopathy;  
KM laser surgery scarring; glaucoma filtering surgery scarring;  
KM macular degeneration; CRST syndrome; bacterial infection; viral disease;  
KM solid neoplasm; vascular tumour; lung; colon; prostate; melanoma;  
KM Kaposi's sarcoma; enzyme.  
XX  
XX Homo sapiens.  
OS  
PN US2004001801-A1.  
PD 01-JAN-2004.  
XX  
XX 23-MAY-2002; 2002US-00156214.  
XX  
XX 23-MAY-2002; 2002US-00156214.  
XX  
XX (CORV-) CORVAS INT INC.  
XX



PI Madison EL, Semple JE, Vlasuk GP, Kemp SJ, Komandla M, Siew DV;  
 XX WPI, 2004-190126/18.  
 DR N-PSDB; ADJ46894.  
 XX Conjugate useful for treating e.g. cancer, cell-surface protease-  
 PT associated diseases, comprising a peptidic substrate or nucleic acid  
 PT substrate linked to a therapeutic agent through a linker.  
 XX  
 PS Claim 9, SEQ ID NO 2, 361bp; English.

XX The invention relates to a conjugate comprising a therapeutic agent and a  
 CC peptidic substrate or nucleic acid substrate linked to the agent  
 CC optionally through a linker or peptidic linker, where the peptidic  
 CC substrate is proteolytically cleaved by a cell surface protease or a  
 CC soluble, released or shed form conjugate to liberate the agent and the  
 CC conjugate is not substantially cleaved by plasmin or prostate specific  
 CC antigen (PSA). The conjugate is useful for treating a disease, which  
 CC involves administering a conjugate to a subject, where the disease is  
 CC preferably a proliferative disease or a cell-surface protease-associated  
 CC disease. The diseases include autoimmune diseases, inflammatory diseases,  
 CC infectious diseases and endocrine diseases. The conjugate is useful for  
 CC treating a cell-surface protease-associated disease, which involves  
 CC administering a conjugate comprising an agent and a peptidic substrate to  
 CC a subject exhibiting symptoms of a cell-surface protease-associated  
 CC disorder, where the disease is selected from cancer, ocular disorders,  
 CC cardiovascular disorders, chronic inflammatory diseases, wounds,  
 CC circulatory disorders, dermatological disorders, restenosis, rheumatoid  
 CC arthritis, psoriasis, diabetic retinopathies, scarring from laser  
 CC surgery, scarring from glaucoma filtering surgery, macular degeneration,  
 CC CRST syndrome, bacterial infections, viral diseases, solid neoplasms and  
 CC vascular tumors such as lung cancer, colon cancer, prostate cancer,  
 CC melanoma and Kaposi's sarcoma. This sequence represents a transmembrane  
 CC serine protease (Misp) polypeptide of the invention.  
 XX  
 SQ Sequence 855 AA;

Query March 99.4%; Score 4653; DB 8; Length 855;  
 Beat Local Similarity 99.5%; Pred. No. 2.4e-307;

Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDRARKGGCGPDPFGAGLKYNSRHEKNGLEGEVEFLPVNNVYKKEKHPGHWVLA 60  
 Db 1 MGSDRARKGGCGPDPFGAGLKYNSRHEKNGLEGEVEFLPVNNVYKKEKHPGHWVLA 60  
 QY 61 VLIGLLVLLGIGFLVHLYQYRDYRVQKYNKYNKYNKYNKYNKYNKYNKYNKYNKYN 120  
 Db 61 VLIGLLVLLGIGFLVHLYQYRDYRVQKYNKYNKYNKYNKYNKYNKYNKYNKYNKYN 120  
 QY 121 KDALKLYSGVPLGPRHYKESAVTAFAESGVIAYYWSFSLPOHLYEAEVMAEERVM 180  
 Db 121 KDALKLYSGVPLGPRHYKESAVTAFAESGVIAYYWSFSLPOHLYEAEVMAEERVM 180  
 QY 181 LPPARSLKSFVYTSVVAFPDTSKTVQRTQDNCSCFGLAHAGVELMRETTGPPDPSPYA 240  
 Db 181 LPPARSLKSFVYTSVVAFPDTSKTVQRTQDNCSCFGLAHAGVELMRETTGPPDPSPYA 240  
 QY 241 HARCQMALRGDADSVLSTFRSPDLASCDERGSDLVTYNTLSPMEPHALVOLCGTYP 300  
 Db 241 HARCQMALRGDADSVLSTFRSPDLASCDERGSDLVTYNTLSPMEPHALVOLCGTYP 300  
 QY 301 YNLTFHSSQNVLLTLITNTERRRHGPATFPQLPRMSSCGRLRKAGQTNSPYYPGHY 360  
 Db 301 YNLTFHSSQNVLLTLITNTERRRHGPATFPQLPRMSSCGRLRKAGQTNSPYYPGHY 360  
 QY 361 PPNIDCTNINIEVNNQHVAFKFFYLLEPGVPAATCKDVEINGEKYCGERSQFVYTS 420  
 Db 361 PPNIDCTNINIEVNNQHVAFKFFYLLEPGVPAATCKDVEINGEKYCGERSQFVYTS 420  
 QY 421 NSNKITVRFHSDQSTYDTGFLAEVLYSYDSPPCPOQFTCRGRCIRKELRCDGWAADCTDH 480  
 Db 421 NSNKITVRFHSDQSTYDTGFLAEVLYSYDSPPCPOQFTCRGRCIRKELRCDGWAADCTDH 480

QY 481 SDELNCSGDAHQFTCKNKECKPLFWVCDVNDGNSDBQSCSPAQTRFCSNGKCLSK 540  
 Db 481 SDELNCSGDAHQFTCKNKECKPLFWVCDVNDGNSDBQSCSPAQTRFCSNGKCLSK 540  
 QY 541 SQQNGKXDCGDSGDEASCPKRVNVVCTKHYRCLNGLCLSKNPECDDKEDCSGSD 600  
 Db 541 SQQNGKXDCGDSGDEASCPKRVNVVCTKHYRCLNGLCLSKNPECDDKEDCSGSD 600  
 QY 601 DCDGCLASFTRQAAVVGCTADDEGEMPOVSLHAGGHI CGASLISPMNLVSAHGYID 660  
 Db 601 DCDGCLASFTRQAAVVGCTADDEGEMPOVSLHAGGHI CGASLISPMNLVSAHGYID 660  
 QY 661 DRGRYSDPTQWTFALIHQDSQASAPGVERRLKRIISHPPDFTFDYDIALLEK 720  
 Db 661 DRGRYSDPTQWTFALIHQDSQASAPGVERRLKRIISHPPDFTFDYDIALLEK 720  
 QY 721 AEYSWVRPCLPDASHVFPAGKAIWVTGHTQYGGTALILQKEIRVINOTTG 780  
 Db 721 AEYSWVRPCLPDASHVFPAGKAIWVTGHTQYGGTALILQKEIRVINOTTG 780  
 QY 781 PQQITPRMVCVPLSGVDSQGGSGPLSSVEADGRI FQAGVWSGDCAGRRKPGVYT 840  
 Db 781 PQQITPRMVCVPLSGVDSQGGSGPLSSVEADGRI FQAGVWSGDCAGRRKPGVYT 840  
 QY 841 RLPLFRDMIKENTGV 855  
 Db 841 RLPLFRDMIKENTGV 855

RESULT 27  
 ID ADO55145  
 ADO55145 standard; protein; 853 AA.

AC ADO55145;  
 DT 15-JUN-2004 (first entry)  
 XX

DE Protein #47 with increased gene expression in renal cell carcinoma.

XX cyrostatic; gene therapy; differential expression; renal cell carcinoma;  
 KW clear cell RCC; papillary RCC; chromophobe/oncocytoma RCC;  
 KW sarcomatoid RCC; RCC; Wilms' tumor; gene expression; kidney cancer;  
 KW diagnostic marker; cancer.

OS Homo sapiens.

PN W02004032842-A2.

PD 22-APR-2004.

PE 06-OCT-2003; 2003WO-US031476.

PR 04-OCT-2002; 2002US-0415775P.

PA (VAND-) VAN ANDEL INST.

PI Teh BT, Takahashi M;

XX WPI, 2004-340789/31.

DR N-PSDB; ADO54959.

XX New nucleic acid and polypeptide compositions, useful in the field of  
 PT molecular biology and medicine, in particular for gene expression  
 PT profiling, identifying diagnostic markers, and treating certain types of  
 PT kidney cancer.

PS Example IV; SEQ ID NO 242; 53bp; English.

CC The invention relates to novel genes that are differentially expressed in  
 CC sub-types of renal cell carcinomas and methods of detecting them using  
 CC nucleic acids and probes. The nucleic acid probes hybridize with part or  
 CC all of a coding sequence that is overexpressed in clear cell renal cell  
 CC carcinoma (CC-RCC), papillary RCC, chromophobe/oncocytoma RCC,

CC sarcomatoid RCC, TCC, or Wilms' tumors, which overexpression is based on  
CC comparison to a baseline value. The methods and compositions of the  
CC present invention are useful in the field of molecular biology and  
CC medicine, in particular for gene expression profiling of certain types of  
CC kidney cancer, in identifying diagnostic markers, and treating such  
CC cancer patients. This sequence corresponds to the protein encoded by a  
CC gene with increased expression in CC-RCC.

XX Sequence 853 AA:

Query Match 98.8%; Score 4625; DB 8; Length 853;

Best Local Similarity 99.4%; Pred. No. 1.9e-305;

Matches 850; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 MGSDRAKGGGGPGDGFAGLKYNSRHRKNGLEGEVFLPVNNVKKYKPGKRWVLA 60  
DB 1 MGSDRAKGGGGPGDGFAGLKYNSRHRKNGLEGEVFLPVNNVKKYKPGKRWVLA 60  
QY 61 VLIIGLLVLIIGLIGFLVWHLQYRDVYQKYNKGYMRITNENFVAYENSNSTEFVSLASKV 120  
DB 61 VLIIGLLVLIIGLIGFLVWHLQYRDVYQKYNKGYMRITNENFVAYENSNSTEFVSLASKV 120  
QY 121 KDALKILYSGVPLGPHKESAVTAABEGSVIATYNEFSFPGHLYEAEARVMAEERVVM 180  
DB 121 KDALKILYSGVPLGPHKESAVTAABEGSVIATYNEFSFPGHLYEAEARVMAEERVVM 180  
QY 181 LPPRARSLSFVWTSVVAAPFDSKTVOITDONSFCFLHARGVYLMFTTGGFSDSPYPA 240  
DB 181 LPPRARSLSFVWTSVVAAPFDSKTVOITDONSFCFLHARGVYLMFTTGGFSDSPYPA 240  
QY 241 HARCQMLRGDADSVLSLTFRSFDLASCDERGSDLVTYNTLSFMEPHALVOLGTYPPS 300  
DB 240 HA-CQMLRGDADSVLSLTFRSFDLASCDERGSDLVTYNTLSFMEPHALVOLGTYPPS 298  
QY 301 YNLFHSSQVNLITLLTNTERRHHPGEATFPQLPRMSSCGGRIRKAQGTFSNPPYRGHY 360  
DB 299 YNLFHSSQVNLITLLTNTERRHHPGEATFPQLPRMSSCGGRIRKAQGTFSNPPYRGHY 358  
QY 361 PPNIDCTWNIIEVPNNQHVAFKFFYLLERPGVAGTCKPKDVEINGEKYCGERSQFVWTS 420  
DB 359 PPNIDCTWNIIEVPNNQHVAFKFFYLLERPGVAGTCKPKDVEINGEKYCGERSQFVWTS 418  
QY 421 NSNKITVRFHSDQSYDTGFLAEYLSYDSDPCGQFTCIRTGRCIRKELRCGNADCTDH 480  
DB 419 NSNKITVRFHSDQSYDTGFLAEYLSYDSDPCGQFTCIRTGRCIRKELRCGNADCTDH 478  
QY 481 SDELNCSGDAGHOTCTCNKFCCKPLFWCDVNDGDNSSDEQSCCPAQTRFCSNGKCLSK 540  
DB 479 SDELNCSGDAGHOTCTCNKFCCKPLFWCDVNDGDNSSDEQSCCPAQTRFCSNGKCLSK 538  
QY 541 SQQNGSGDDCGDSDDEASCPKVNVTCTKHTYRCLANGCLSKNGPECDGKEDCSDGSD 600  
DB 539 SQQNGSGDDCGDSDDEASCPKVNVTCTKHTYRCLANGCLSKNGPECDGKEDCSDGSD 598  
QY 601 DCDGRLSFRQARVVGSTDADEGEPMQVSLHALGQSHICGASLISPNMLVSAHACYID 660  
DB 599 DCDGRLSFRQARVVGSTDADEGEPMQVSLHALGQSHICGASLISPNMLVSAHACYID 658  
QY 661 DRGRYSDDPQWMTAFILHDOSQHSABGVQERRIKRIISHPFFNDPFDDIALLELEKP 720  
DB 659 DRGRYSDDPQWMTAFILHDOSQHSABGVQERRIKRIISHPFFNDPFDDIALLELEKP 718  
QY 721 AEYSMWRTICLPASHVFPAGKAIWYGMCHTOYGGTALILQKGIKRVINOTTGNNL 780  
DB 719 AEYSMWRTICLPASHVFPAGKAIWYGMCHTOYGGTALILQKGIKRVINOTTGNNL 778  
QY 781 PQQITPRMWCVGLISGVDSQGDGSGPLSSVEADGRIFGAGVVSWDGCGAGRNKPGVYT 840  
DB 779 PQQITPRMWCVGLISGVDSQGDGSGPLSSVEADGRIFGAGVVSWDGCGAGRNKPGVYT 838  
QY 841 RLPLFRDWIKENTGV 855  
DB 839 RLPLFRDWIKENTGV 853

RESULT 28

AAM25628

ID AAM25628 standard; protein; 851 AA.

XX AAM25628;

DT 16-OCT-2001 (first entry)

DE Human protein sequence SEQ ID NO:1143.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
XX antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
XX antibacterial; endocrine; cardiant; central nervous system; virucide;  
XX anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
XX antiaggregant; haemostatic; vulnery; antilucer; osteopathic; eczema;  
XX dermatological; antiallergic; antistimatic; antidiabetic; cyrostatic;  
XX neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
XX immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
XX antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
XX cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
XX genetic disease; haematopoietic disorder; platelet disorder; asthma;  
XX thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
XX allergic rhinitis; diabetes; multiple sclerosis; depression;  
XX Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
XX neurological disorder.

OS Homo sapiens.

PN WO200153455-A2.

PD 26-JUL-2001.

PF 22-DEC-2000; 2000MO-US035017.

XX 23-DEC-1999; 99US-00471275.

XX 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-457603/49.

XX N-SDB; AAM99569.

XX Isolated human polynucleotides encoding polypeptides, useful for the

XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection.

XX Claim 20; Page 237; 1217pp; English.

XX AAM99166 to AAM99904 encode the human proteins given in AAM25225 to

XX AAM25863. The proteins can have activities based on the tissues and cells

XX they are expressed in, such as: antiinflammatory; antirheumatic;

XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;

XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;

XX cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;

XX antilucer; osteopathic; dermatological; antiallergic; antistimatic;

XX antidiabetic; cyrostatic; neuroprotective; antidepressant; nootropic;

XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides

XX encoding them can be used in gene therapy, antisense therapy and vaccine

XX production. The proteins and polynucleotides are useful for screening for

XX agonists or antagonists of a protein and for the treatment and diagnosis

XX of disorders associated with the activity of a protein e.g. inflammation,

XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,

XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal

XX infections, autoimmunity, genetic diseases, haematopoietic disorders,

XX anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,

XX osteoporosis, severe combined immunodeficiency, eczema, allergic

XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,

XX Alzheimer's disease, Parkinson's disease, neurodegenerative and

XX neurological disorders

XX	Sequence 851 AA:		
90	Query Match	98.5%; Score 4614; DB 4; Length 851;	
	Best Local Similarity	99.3%; Pred. No. 1,1e-304;	
	Matches 843; Conservative 1; Mismatches 5; Indels 0; Gaps 0;		
QY	7	RKGGGGKPDPAAGLKYNRRHKNVGLBEGVFLLPVNNVKKYKKGPGRWVLAIVLIGLL	66
DB	3	RKGGGGKPDPAAGLKYNRRHKNVGLBEGVFLLPVNNVKKYKKGPGRWVLAIVLIGLL	62
QY	67	LVLIGIFLWHLQYRDVAVQKYNKGYRINENFVDAEYNSNTEFVSLASKYDALKL	126
DB	63	LVLIGIFLWHLQYRDVAVQKYNKGYRINENFVDAEYNSNTEFVSLASKYDALKL	122
QY	127	LYSGVPLGPIYHKSAAVTAFFSEGSVIAVYWFSEFSIPQHLVEABERVMAEERVWLPPRAR	186
DB	123	LYSGVPLGPIYHKSAAVTAFFSEGSVIAVYWFSEFSIPQHLVEABERVMAEERVWLPPRAR	182
QY	187	SLKSFVVTSVVAFPTDSKTIVORTDNGSCFGLHARGVELMFTTGPDPSPYPHARQW	246
DB	183	SLKSFVVTSVVAFPTDSKTIVORTDNGSCFGLHARGVELMFTTGPDPSPYPHARQW	242
QY	247	ALRGDADSVLSLTFRSFPLASCDERGSPLVTVYNTLSFMEPHALVOLCGTTPSPYNLTFH	306
DB	243	ALRGDADSVLSLTFRSFPLASCDERGRHLVTVYNTLSFMEPHALVOLCGTTPSPYNLTFH	302
QY	307	SSQNVLLITLTNTERRHPPGEATFPLPRMSSCGGLRKAQGTFSNPPYPGHYPPNIDC	366
DB	303	SSQNVLLITLTNTERRHPPGEATFPLPRMSSCGGLRKAQGTFSNPPYPGHYPPNIDC	362
QY	367	TWNIIEVNNQVKKRFRFFYLLEPGVPAGTCKPYVEINSGKYCGESQFVYVTSNKKIT	426
DB	363	TWNIIEVNNQVKKRFRFFYLLEPGVPAGTCKPYVEINSGKYCGESQFVYVTSNKKIT	422
QY	427	VRFISDQSYDTGTGLAEYLSYDSDPCPGQFTCRGRCIRRELRCDDGADCTDSDELNC	486
DB	423	VRFISDQSYDTGTGLAEYLSYDSDPCPGQFTCRGRCIRRELRCDDGADCTDSDELNC	482
QY	487	SCDAGHOFCKNKECKPLFWVCDSDVNDGSDSDGSCSPAQTFRCNSGKCLSKSQQNG	546
DB	483	SCDAGHOFCKNKECKPLFWVCDSDVNDGSDSDGSCSPAQTFRCNSGKCLSKSQQNG	542
QY	547	KDDCGDSDASCRNVVVTCTKITYRCLNGLCLSKNPEBCDGGKDCSDSDDEKDCDGL	606
DB	543	KDDCGDSDASCRNVVVTCTKITYRCLNGLCLSKNPEBCDGGKDCSDSDDEKDCDGL	602
QY	607	RSPTROARVVGTDADDEGEMPMOYSLHLAGQGHICGASLISPNMLVSAHNCYIDRGRY	666
DB	603	RSPTROARVVGTDADDEGEMPMOYSLHLAGQGHICGASLISPNMLVSAHNCYIDRGRY	662
QY	667	SDPTQMTAFGLHDSQSRAPGVERRLKRIISHPFNDFTFDYDIALLELEKPAEYSSM	726
DB	663	SDPTQMTAFGLHDSQSRAPGVERRLKRIISHPFNDFTFDYDIALLELEKPAEYSSM	722
QY	727	VRIPTCLPDASHVFPAGKAIWTTGNGHTYGGTGALLDKGIRIYNQTTCNLLPQQTTP	786
DB	723	VRIPTCLPDASHVFPAGKAIWTTGNGHTYGGTGALLDKGIRIYNQTTCNLLPQQTTP	782
QY	787	RMMCVCGLSGGVDSQGGSGGPLSSVEADGRIFGAGVSVWMDGCGRRKPPVYTRLPFLR	846
DB	783	RMMCVCGLSGGVDSQGGSGGPLSSVEADGRIFGAGVSVWMDGCGRRKPPVYTRLPFLR	842
QY	847	DWIKENTGV 855	
DB	843	DWIKENTGV 851	

XX	11-JAN-2002 (first entry)		
DE	Human membrane-type Ser kinase homologue, SEQ ID NO:1798.		
XX	Human; cytokine; cell proliferation; cell differentiation; growth factor;		
XX	haematopoietic regulation; tissue growth; immunomodulator; actinin;		
KM	inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;		
KM	proliferation; metastasis; cancer; tumour; haematopoietic disorder;		
KM	myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;		
KM	chronic inflammatory condition; proliferative retinopathy;		
KM	atherosclerosis; coronary heart disease; arterial ischaemia;		
KM	bone disorder; osteoporosis; vascular growth disorder;		
KM	tissue regeneration; wound healing; infection; immune disorder;		
KM	cell culture; drug screening; gene therapy; antiinflammatory;		
KM	antiaesthetic; antiarthritic; haemostatic; antiarteriosclerotic;		
KM	cytotoxic; osteopathic; vasotropic; cardiant; virucide; antibacterial;		
KM	antifungal; vulnery; antitumor.		
XX			
OS	Homo sapiens.		
XX			
XX	MO200157188-A2.		
XX			
PD	09-AUG-2001.		
XX			
PE	05-FEB-2001; 2001MO-US003800.		
XX			
PR	03-FEB-2000; 2000US-00496914.		
XX			
PR	27-APR-2000; 2000US-00560875.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Drmanac RT;		
XX			
DR	WPI; 2001-457740/49.		
XX			
DR	N-PSDB; ABA08672.		
PT	Human proteins and DNA encoding sequences useful for preventing, treating		
PT	or ameliorating a medical condition in a mammalian subject e.g. arthritis		
PT	and cancer.		
XX			
PS	Claim 20; Page 188; 1963pp; English.		
XX			
CC	Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and		
CC	sequences ABA08225-ABA09574 represent nucleic acids encoding them. The		
CC	invention also relates to vectors and recombinant host cells comprising a		
CC	nucleotide of the invention, methods of producing the novel polypeptides,		
CC	antibodies against the polypeptides, methods of detecting the nucleotides		
CC	or polypeptides in a sample, and methods of identifying compounds which		
CC	bind to polypeptides of the invention. Although novel, many of the		
CC	polypeptides of the invention have homology to known proteins, thereby		
CC	giving an insight into their probable biological activities, and hence		
CC	potential therapeutic applications. The polypeptides of the invention may		
CC	have various activities, including cytokine, cell proliferation or cell		
CC	differentiation activities; stem cell growth factor activity;		
CC	haematopoietic regulatory activity; tissue growth activity;		
CC	immunomodulatory activity; activin- or inhibin-related activities;		
CC	chemotactic or chemokinetic activities; haemostatic, thrombotic or		
CC	thrombolytic activities; receptor or ligand activities; or may be		
CC	involved in oncogenesis, cancer cell proliferation or metastasis.		
CC	Depending on their biological activities, polypeptides and nucleotides of		
CC	the invention are useful for preventing, treating or ameliorating medical		
CC	conditions, e.g., by protein or gene therapy. Such conditions include		
CC	cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell		
CC	disorders), chronic inflammatory conditions (e.g., asthma or arthritis),		
CC	proliferative retinopathy, atherosclerosis, coronary heart disease,		
CC	arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal		
CC	vascular growth. Polypeptides involved with tissue regeneration and		
CC	repair (or nucleic acids encoding them) may be used to promote wound		
CC	healing (e.g., of burns, incisions and ulcers), while those with		
CC	immunomodulatory activities may be used in the treatment of viral,		
CC	bacterial and fungal infections in addition to immune disorders.		
CC	Polypeptides with growth factor activity may be used in cell cultures to		

CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a novel human  
CC polypeptide of the invention  
XX  
XX

SQ Sequence 851 AA;

Query Match 98.5%; Score 4614; DB 4; Length 851;

Best Local Similarity 99.3%; Pred. No. 1.1e-304; Mismatches 5; Indels 0; Gaps 0;

Matches 843; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```
QY 7 RKGGGGFKDFGAGLKYNSRHEKVNGLBEGVEFLPVNNVKKVEKKGPRWVLAAILGLL 66
DB 3 RKGGGGFKDFGAGLKYNSRHEKVNGLBEGVEFLPVNNVKKVEKKGPRWVLAAILGLL 62
QY 67 LVLLIGFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSNSTEFVSLASKYKDALKL 126
DB 63 LVLLIGFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSNSTEFVSLASKYKDALKL 122
QY 127 LVGGVPELGPYHKEASATAFSEGSVIAVYVSESEIPQHLVBEARVMAEERVWMLPPRAR 166
DB 123 LVGGVPELGPYHKEASATAFSEGSVIAVYVSESEIPQHLVBEARVMAEERVWMLPPRAR 182
QY 187 SLKSFVTVSVAAPEPTDSKTVQRTQDNSCFGLHARGVLELRFITPGFPPSPYAHARQW 246
DB 183 SLKSFVTVSVAAPEPTDSKTVQRTQDNSCFGLHARGVLELRFITPGFPPSPYAHARQW 242
QY 247 ALRGDADSVLSLTFRSDLASCDERGSDLVTVNTLSPMERHALVOLCGTYPSPSYNTTFH 306
DB 243 ALRGDADSVLSLTFRSDLASCDERGRHLTVVNTLSPMERHALVOLCGTYPSPSYNTTFH 302
QY 307 SSONVLLITLITINTERHPCGPBATEFQLPMSSCGRLRAQGFENSPPYRGHYPVYDC 366
DB 303 SSONVLLITLITINTERHPCGPBATEFQLPMSSCGRLRAQGFENSPPYRGHYPVYDC 362
QY 367 TWINIEVNNOHVKVRFEFYLLBEPVAGTCPCPDYVEINSEKCYGERSQEPVTSNSNKT 426
DB 363 TWINIEVNNOHVKVRFEFYLLBEPVAGTCPCPDYVEINSEKCYGERSQEPVTSNSNKT 422
QY 427 VRHSDSDSYDTGFLAFLYLSYSDSDPCPGQFTCTGRCIRKEARCDGMADCTHSDBLNC 486
DB 423 VRHSDSDSYDTGFLAFLYLSYSDSDPCPGQFTCTGRCIRKEARCDGMADCTHSDBLNC 482
QY 487 SCDAQHOFCKNRCCKPLFWVCDSDVNDGNSDDEGSCPAQPFRCGNSGKCLSKSQCCNG 546
DB 483 SCDAQHOFCKNRCCKPLFWVCDSDVNDGNSDDEGSCPAQPFRCGNSGKCLSKSQCCNG 542
QY 547 KDDCGDSDASCPKXNVVTCTKHTYRCLNGLCLSKGNPECDGKEDCDSDDEKDCDGL 606
DB 543 KDDCGDSDASCPKXNVVTCTKHTYRCLNGLCLSKGNPECDGKEDCDSDDEKDCDGL 602
QY 607 RSTFRQARVVGCTDADGEMPMQVSLALQCGHICGASLISPMVLVAACIYIDDRGFY 666
DB 603 RSTFRQARVVGCTDADGEMPMQVSLALQCGHICGASLISPMVLVAACIYIDDRGFY 662
QY 667 SDPTQWAFPLGLHDQSORSAFGVQERLKRILISHPFNDTFPPVDIALLELEKPAEYSSM 726
DB 663 SDPTQWAFPLGLHDQSORSAFGVQERLKRILISHPFNDTFPPVDIALLELEKPAEYSSM 722
QY 727 VRPILCPDASHVPBAGKAIWVTGHTQYGGTGLILQKEIRVINTCTGENTLPOQITP 786
DB 723 VRPILCPDASHVPBAGKAIWVTGHTQYGGTGLILQKEIRVINTCTGENTLPOQITP 782
QY 787 RMMCVGFLSGGVDSCCQDSSGGLSSVADGRIFGAGVVSNGDGCAGNNKGVYTRLPFLR 846
DB 783 RMMCVGFLSGGVDSCCQDSSGGLSSVADGRIFGAGVVSNGDGCAGNNKGVYTRLPFLR 842
QY 847 DWIKENTGV 855
DB 843 DWIKENTGV 851
```

RESULT 30

ABG21442

ID ABG21442 standard; protein; 932 AA.

XX ABG21442;

XX 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #21433.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN NC0200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

PI Dermanac RT, Liu C, Tang YT;

XX WPI, 2001-639362/73.

DR N-PSDB; AAS85629.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX Claim 20; SEQ ID NO 51801; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX sequences. (I) is useful as hybridisation probes, polymerase chain

XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX and in recombinant production of (II). The polynucleotides are also used

XX in diagnostics as expressed sequence tags for identifying expressed

XX genes. (I) is useful in gene therapy techniques to restore normal

XX activity of (II) or to treat disease states involving (II). (II) is

XX useful for generating antibodies against it, detecting or quantitating a

XX polypeptide in tissue, as molecular weight markers and as a food

XX supplement. (II) and its binding partners are useful in medical imaging

XX of sites expressing (II). (I) and (II) are useful for treating disorders

XX involving aberrant protein expression or biological activity. The

XX polypeptide and polynucleotide sequences have applications in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits to assess biodiversity

XX and to produce other types of data and products dependent on DNA and

XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

XX patent did not appear in the printed specification, but was obtained in

XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 932 AA;

Query Match 91.9%; Score 4302; DB 4; Length 932;

Best Local Similarity 95.8%; Pred. No. 1.9e-283; Mismatches 24; Indels 4; Gaps 4;

Matches 796; Conservative 7; Mismatches 24; Indels 4; Gaps 4;

```
QY 27 EKVNGLEBEGVEFLPVNNVKKVEKKGPRWVLAAILGLLVLIGLIGFLVWHLQYRDVRY 86
DB 100 QKVNGLEBEGVEFLPVNNVKKVEKKGPRWVLAAILGLLVLIGLIGFLVWHLQYRDVRY 159
QY 87 QKVNGYMRITNENFVDAYENSNSTEFVSLASKYKDALKLIVSGVPELGPYHKEASATAF 146
```

Db 160 QKVNGVIRINENFVDAYENSNTSEFVSLASKVDALKLYSGVPLGPHKESAATAF 219  
QY 147 SEGSVIAYYWESEFSI PQHLVEAEERVAEERVVMLEPPARSLKSFVTVSVVAFPTDSKTV 206  
Db 220 SEGSVIAYYWESEFSI PQHLVEAEERVAEERVVMLEPPARSLKSFVTVSVVAFPTDSKTV 279  
QY 207 QRTODNSCSFGLHARGVELMRFTTPGFPDSEYPAPARCOMALRGDADSVLSLFRSPDLA 266  
Db 280 QRTODNSCSFGLHARGVELMRFTTPGFPDSEYPAPARCOMALRGDADSVLSLFRSPDLA 339  
QY 267 SCDERGSDLVTYNTLSPEMPEHALVOLCGTYPSPSYNLTFHSSQVNLITLITNTERRHFG 326  
Db 340 SCDERGRILVTYNTLSPEMPEHALVOLCGTYPSPSYNLTFHSSQVNLITLITNTERRHFG 399  
QY 327 FEATFPQLPRMSSCGGRLRKAQGTFNSPYYPGHYRP-NIDCTWNI EVPNNOHVKKRFKFP 385  
Db 400 FEATFPQLPRMSSCGGRLRKAQGTFNSPYYPGHYRPQHXSTWNIEVPNNOHVKKRFKFP 459  
QY 386 YLLERGVAGTCPRDYVINEKEYCGERSQFVTVSNKTIIVRFHSDQSYTDGFLAEYL 445  
Db 460 YLLERGVAGTCPRDYVINEKEYCGERSQFVTVSNKTIIVRFHSDQSYTDGFLAEYL 519  
QY 446 SYDSSDPCPGQFTCRTR-CIRKELRCDMADCTDHSDELNCSCDAGHQFTCKNFKCKPL 504  
Db 520 SYDSSDPCPGQFTCRTRCIRKELRCDMADCTDHSDELNCSCDAGHQFTCKNFKCKPL 579  
QY 505 FMVCDSVNDGSDNDEGSCSPAGTFRCSNGKCLSKSQCNKGKDDCGDSDASCPKXNV 564  
Db 580 FMVCDSVNECGDNDDEGSCIPAGTFRCSNGKCLSKSQCNKGKDDCGDSDASCPKXNV 639  
QY 565 VTCTKHTYRCINGLCLSKGNPECDGKEDCSGSDKDCDCLASFTROARVVGTTDADEG 624  
Db 640 VTCTKHTYRCINGLCLSKGNPECDGKEDCSGSDKDCDCLASFTROARVVGTTDADEG 699  
QY 625 EPMQVSLHALGQGHICGASLISPMVLVSAHCVIIDRGFRYSDPTQWT-AFLGLHDQSQ 683  
Db 700 EPMQVSLHALGQGHICGASLISPMVLVSAHCVIIDRRFMYSNPTQWNRRAFLGLHDQSQ 759  
QY 684 RSAP-GVQERRLKRIISHPFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAG 742  
Db 760 RRPWGCGRKRLKRIISHPFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAG 819  
QY 743 KAIWVTGKHTOYGGTGALLIOKGEIRVINOITTCENLLPQOITPRMVCVGLSGVDSQ 802  
Db 820 KAIWVTGKHTOYGGTGALLIOKGEIRVINOITTCENLLPQOITPRMVCVGLSGVDSQ 879  
QY 803 GDSGGLSSVEADGRIFGAGVVMGDCAGRNKPGVYTRLLPLFRDMIKENTGV 855  
Db 880 GDSGGLSSVEADGRIFGAGVVMGDCAGRNKPGVYTRLLPLFRDMIKENTGV 932

Search completed: November 29, 2004, 08:28:11  
Job time : 129.858 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: November 29, 2004, 08:19:13 : Search time 142.315 Seconds

(without alignments)  
3456.743 Million cell updates/sec

Title: US-09-936-333-27

Perfect score: 4683  
Sequence: 1 MGSBRARKGCGGPGKDFAGL.....RGVYTRLPFRDKWIKENTGV 855Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database :

1: uniprot\_02:.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4659	99.5	855	1	ST14_HUMAN
2	3884	82.9	855	1	ST14_MOUSE
3	3884	82.9	855	1	BAC35125
4	3866	82.6	855	2	Q9J117
5	2686	57.4	845	2	Q6GR54
6	2660	56.8	845	2	Q9DGR1
7	2365	50.5	422	2	Q8WVCI
8	1965	42.0	663	2	Q6DEV0
9	1114.5	23.8	799	2	Q6PF94
10	1114.5	23.8	799	2	AAH57674
11	1114.5	23.8	811	1	TMS6_MOUSE
12	1114	23.8	811	1	TMS6_MOUSE
13	1112	23.7	802	2	Q6UXD6
14	1112	23.7	802	2	AAQ88764
15	1092	23.3	824	2	Q6ICC2
16	1092	23.3	824	2	CAG30332
17	1004.5	21.4	572	2	Q7RTY8
18	1004.5	21.4	572	2	Q8BIK6
19	1004.5	21.4	572	2	BAD18401
20	724	15.5	1034	1	ENTK_PIG
21	711.5	15.2	855	2	Q7Z410
22	711.5	15.2	1059	2	Q7Z411
23	709.5	15.2	1035	1	ENTK_BOVIN
24	695	14.8	1042	1	CORI_HUMAN
25	694.5	14.8	1111	2	Q80YN4
26	684	14.6	777	2	Q8CANG
27	679.5	14.5	1019	1	ENTK_HUMAN
28	673.5	14.4	1069	1	ENTK_MOUSE
29	661.5	14.1	1113	1	CORI_MOUSE
30	639.5	13.7	767	2	Q9DGR2
31	639	13.6	722	2	Q6NUP5

32	639	13.6	722	2	AAH68636	AAH68636 xenopus 1
33	633.5	13.5	680	2	Q868H7	Q868H7 branchiosto
34	616.5	13.2	680	2	Q868H5	Q868H5 branchiosto
35	615	13.1	581	2	Q9XZM7	Q9XZM7 strongyloce
36	609	13.0	688	2	Q868H6	Q868H6 branchiosto
37	596	12.7	490	2	Q7TND4	Q7TND4 mus musculu
38	596	12.7	490	2	Q920K3	Q920K3 rattus norv
39	595	12.7	490	1	TMS2_MOUSE	Q9168 mus musculu
40	594.5	12.7	490	2	Q6P7D7	Q6P7D7 rattus norv
41	594.5	12.7	490	2	AAH61712	AAH61712 rattus norv
42	594	12.7	688	2	Q868H4	Q868H4 branchiosto
43	589.5	12.6	701	2	Q9JUS9	Q9JUS9 rattus norv
44	583	12.4	422	1	DESI_HUMAN	Q9158 homo sapien
45	583	12.4	423	2	Q6UW31	Q6UW31 homo sapien
46	583	12.4	423	2	AAQ89376	AAQ89376 homo sapi
47	583	12.4	703	2	Q8CHN8	Q8CHN8 rattus norv
48	582.5	12.4	704	1	CRAR_MOUSE	P88064 mus musculu
49	581.5	12.4	868	2	Q9Y1V3	Q9Y1V3 polyandroca
50	576.5	12.3	1134	2	Q7RTY7	Q7RTY7 homo sapien
51	576.5	12.3	1524	2	Q91674	Q91674 xenopus lae
52	570	12.2	699	1	CRAR_HUMAN	P48740 h complemen
53	567	12.1	418	2	Q6IEI5	Q6IEI5 rattus norv
54	567	12.1	676	2	Q6DUU6	Q6DUU6 cyprinus ca
55	556	11.9	492	2	Q6ET73	Q6ET73 homo sapien
56	555.5	11.9	492	1	TMS2_HUMAN	O15393 homo sapien
57	555.5	11.9	492	2	AAH51839	AAH51839 homo sapi
58	554	11.8	681	2	Q7ZT70	Q7ZT70 lampetra ja
59	549	11.7	698	2	Q6GPF9	Q6GPF9 xenopus lae
60	548.5	11.7	60	2	Q8BM10	Q8BM10 mus musculu
61	541	11.6	307	2	Q6ZND6	Q6ZND6 homo sapien
62	541	11.6	307	2	BAD18439	BAD18439 homo sapi
63	540	11.5	453	2	Q812N6	Q812N6 mus musculu
64	539	11.5	453	1	TMS3_MOUSE	Q812N6 mus musculu
65	538.5	11.5	417	2	Q8VH04	Q8VH04 rattus norv
66	536.5	11.5	453	2	Q6ZWC3	Q6ZWC3 homo sapien
67	536.5	11.5	453	2	AAQ88823	AAQ88823 homo sapi
68	536.5	11.5	453	2	BAD18806	BAD18806 homo sapi
69	534	11.4	707	2	Q8CGV0	Q8CGV0 cyprinus ca
70	532.5	11.4	279	2	Q9QZ74	Q9QZ74 rattus norv
71	532.5	11.4	638	2	Q8R0P5	Q8R0P5 mus musculu
72	528.5	11.3	618	1	KAL_MOUSE	P26262 mus musculu
73	527.5	11.3	617	2	Q8J181	Q8J181 triaklis scy
74	526	11.2	417	2	TMS3_HUMAN	P57727 homo sapien
75	522.5	11.2	417	2	Q8VDY1	Q8VDY1 mus musculu
76	522.5	11.2	417	2	Q8VHK6	Q8VHK6 mus musculu
77	520.5	11.1	279	2	Q7TNX3	Q7TNX3 mus musculu
78	518.5	11.1	600	2	Q7ZTR2	Q7ZTR2 xenopus lae
79	515	11.0	603	1	CFAI_MOUSE	O61129 mus musculu
80	515	11.0	686	2	Q6OLQ9	Q6OLQ9 gallus gall
81	515	11.0	686	2	AAH73179	AAH73179 gallus ga
82	514.5	11.0	698	2	Q9PU71	Q9PU71 xenopus lae
83	514	11.0	604	1	CFAI_RAT	Q9UWU3 rattus norv
84	510.5	10.9	697	2	Q8CG43	Q8CG43 rattus norv
85	510.5	10.9	733	2	Q8CD27	Q8CD27 mus musculu
86	510.5	10.9	733	2	Q920S0	Q920S0 mus musculu
87	508.5	10.9	418	1	HATT_HUMAN	Q60235 homo sapien
88	508	10.8	455	1	TMS5_MOUSE	Q9ZT04 mus musculu
89	507.5	10.8	638	1	KAL_RAT	P14272 rattus norv
90	506.5	10.8	790	1	PLMN_PIG	P06867 sus scrofa
91	504.5	10.8	643	2	Q6ZWK6	Q6ZWK6 homo sapien
92	504.5	10.8	438	2	BAC85495	BAC85495 homo sapi
93	504.5	10.8	537	2	Q9BYR1	Q9BYR1 homo sapien
94	504.5	10.8	581	2	Q9BIE2	Q9BIE2 homo sapien
95	504	10.8	471	2	Q8CFE0	Q8CFE0 mus musculu
96	504	10.8	615	2	Q03711	Q03711 xenopus lae
97	504	10.8	643	2	Q7PWE4	Q7PWE4 anopheles g
98	504	10.8	728	2	Q96R84	Q96R84 homo sapien
99	503.5	10.8	445	2	Q8CJ17	Q8CJ17 rattus norv
100	502.5	10.7	688	2	Q9PVY4	Q9PVY4 xenopus lae

## ALIGNMENTS



RESULT 1  
ST14 HUMAN STANDARD: PRT: 855 AA.  
ID Q9V5Y6; Q9BS01; Q9H3S0; Q9HB36; Q9HCA3;  
AC 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matrilysin) (Membrane-type serine protease 1) (MT-SP1) (Proteinase) (Serine protease TADG-15)  
DE (Tumor associated differentially-expressed gene-15 protein).  
GN Name-ST14; Synonyms=PRSS14, SNCL19, TADG15;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99303581; PubMed=10373424;  
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;  
RT "Molecular cloning of cDNA for matrilysin, a matrix-degrading serine protease with trypsin-like activity";  
RL J. Biol. Chem. 274:18231-18236(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99432178; PubMed=10500122;  
RA Takeuchi T., Shuman M.A., Craik C.S.;  
RT "Reverse biochemistry: use of macromolecular protease inhibitors to dissect complex biological processes and identify a membrane-type serine protease in epithelial cancer and normal tissue";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Prostate;  
RA Yamaguchi N., Mitsui S.;  
RT "Molecular cloning of a novel transmembrane serine protease expressed in human prostate";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parnley T.H., O'Brien T.J.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Blood, and Muscle;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C., Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalano D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.N., Krzywinski M.I., Skalek U., Small D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [6]  
RP SEQUENCE OF 340-664 FROM N.A.  
RA Cao J., Fan W., Zheng S.;  
RT "Genomic analysis of a novel human serine protease SMC19";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [7]

RP CHARACTERIZATION.  
RC TISSUE=Milk;  
RX MEDLINE=99303582; PubMed=10373425;  
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;  
RT "Purification and characterization of a complex containing matrilysin and a Kunitz-type serine protease inhibitor from human milk";  
RL J. Biol. Chem. 274:18237-18242(1999).  
CC -1- FUNCTION: Degrades extracellular matrix. Proposed to play a role in breast cancer invasion and metastasis. Exhibits trypsin-like activity as defined by cleavage of synthetic substrates with Arg or Lys as the P1 site.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
CC -1- SIMILARITY: Belongs to peptidase family S1.  
CC -1- SIMILARITY: Contains 2 CUB domains.  
CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF118224; AAD42765.2; -;  
CC EMBL; AF133086; AAP00109.1; -;  
CC EMBL; AB030036; BAB20376.1; -;  
CC EMBL; AF057145; AAG15395.1; -;  
CC EMBL; AF005826; AAH05826.2; -;  
CC EMBL; BC030532; AAH03052.1; -;  
CC EMBL; AF283256; AAG13949.1; -;  
CC PDB; 1EAW; X-ray; A/C=615-855.  
CC MEROPS; S01.302; -;  
CC Genew; HGNC:11344; ST14.  
CC MIM; 606797; -;  
CC GO; GO:0005887; C:integral to plasma membrane; TAS.  
CC GO; GO:0008236; F:serine-type peptidase activity; TAS.  
CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.  
CC InterPro; IPR000859; CUB.  
CC InterPro; IPR002172; LDL\_receptor\_A.  
CC InterPro; IPR001254; Peptidase\_S1.  
CC InterPro; IPR001314; Peptidase\_S1A.  
CC InterPro; IPR009003; Pept\_Ser\_Cys.  
CC Pfam; PF00431; CUB; 2.  
CC Pfam; PF00057; Ldl\_recept\_a; 4.  
CC Pfam; PF00089; Trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC PRINTS; PR00261; LDLRECEPTOR.  
CC SMART; SM00042; CUB; 2.  
CC SMART; SM00192; LDLa; 4.  
CC SMART; SM00020; TRYP\_SPC; 1.  
CC PROSITE; PS01180; CUB; 2.  
CC PROSITE; PS01209; LDLa\_1; 2.  
CC PROSITE; PSS0068; LDLa\_2; 4.  
CC PROSITE; PSS0240; TRYPsin\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
CC 3D-structure; Glycoprotein; Hydrolase; Repeat; Serine protease;  
KW Signal-anchor; Transmembrane.  
FT DOMAIN 1 55 Cytoplasmic (Potential).  
FT TRANSMEM 56 76 Signal-anchor for type II membrane protein (Potential).  
FT DOMAIN 77 855 Extracellular (Potential).  
FT DOMAIN 214 334 CUB 1.  
FT DOMAIN 340 447 CUB 2.  
FT DOMAIN 452 487 LDL-receptor class A 1.  
FT DOMAIN 487 524 LDL-receptor class A 2.  
FT DOMAIN 524 560 LDL-receptor class A 3.  
FT DOMAIN 566 603 LDL-receptor class A 4.  
FT DOMAIN 615 854 Serine protease.  
FT ACT\_SITE 656 Charge relay system (By similarity).  
FT ACT\_SITE 711 Charge relay system (By similarity).  
FT ACT\_SITE 805 Charge relay system (By similarity).  
FT ACT\_SITE 805

FT CARBOHYD 109 109 N-linked (GlcNAc... ) (Potential) .  
 FT CARBOHYD 302 302 N-linked (GlcNAc... ) (Potential) .  
 FT CARBOHYD 485 485 N-linked (GlcNAc... ) (Potential) .  
 FT CARBOHYD 772 772 N-linked (GlcNAc... ) (Potential) .  
 FT CONFLICT 381 381 R -> S (in Ref. 4) .  
 FT CONFLICT 674 674 A -> V (in Ref. 3) .  
 SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;

Query Match 99.5%; Score 4659; DB 1; Length 855;

Best Local Similarity 99.6%; Pred. No. 1e-313; Mismatches 3; Indels 0; Gaps 0;

Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGPGDGFAGLKYNSRHEKVNGLKEGVELEPVNNYKYEKGPHGKRVVLA 60  
 DB 1 MGSDRARKGGGPGDGFAGLKYNSRHEKVNGLKEGVELEPVNNYKYEKGPHGKRVVLA 60  
 QY 61 VLIGLLVLIGLIGFLVWHLQYRDVQVKXKGYRITNENFVDA YENSNSTEYVSLASKV 120  
 DB 61 VLIGLLVLIGLIGFLVWHLQYRDVQVKXKGYRITNENFVDA YENSNSTEYVSLASKV 120  
 QY 121 KDALKLYSGVPLGPHKESAVTAFSGSIATYMSFSIPQHLVEAEKVAEERVYM 180  
 DB 121 KDALKLYSGVPLGPHKESAVTAFSGSIATYMSFSIPQHLVEAEKVAEERVYM 180  
 QY 181 LPPRARSLSKFSVTSVVAFTDSKTVORTODNSCSFGLHARGVELMRPTTGPPDSPYPA 240  
 DB 181 LPPRARSLSKFSVTSVVAFTDSKTVORTODNSCSFGLHARGVELMRPTTGPPDSPYPA 240  
 QY 241 HARCOMALRGDADSVLSLTFRSPDLASCDEBGSJLVTVYNTLSPEMHVQLGTYPPS 300  
 DB 241 HARCOMALRGDADSVLSLTFRSPDLASCDEBGSJLVTVYNTLSPEMHVQLGTYPPS 300  
 QY 301 YNLTFFHSQNVLLTLITNTRRRHPGPAETFOUPRMSCCGRRLRKAQGTNSYRYGHY 360  
 DB 301 YNLTFFHSQNVLLTLITNTRRRHPGPAETFOUPRMSCCGRRLRKAQGTNSYRYGHY 360  
 QY 361 PNPIDCWNINIEVPNNQHVKAFFKFFYLLEPGVPAATCKDVEVINGEKYCEBESQFVYTS 420  
 DB 361 PNPIDCWNINIEVPNNQHVKAFFKFFYLLEPGVPAATCKDVEVINGEKYCEBESQFVYTS 420  
 QY 421 NSNKITVRPHSDQSYDTGTFLAELYSYSDSPCEQFTCRGTGRCIRKELRCGMAADCTDH 480  
 DB 421 NSNKITVRPHSDQSYDTGTFLAELYSYSDSPCEQFTCRGTGRCIRKELRCGMAADCTDH 480  
 QY 481 SDELINCSGDAGHOTCTCKRKFCKPLFWYCDVNDGDNSEBGCSCPAQTFRCSNGKCLSK 540  
 DB 481 SDELINCSGDAGHOTCTCKRKFCKPLFWYCDVNDGDNSEBGCSCPAQTFRCSNGKCLSK 540  
 QY 541 SQQCNKGDDCGDSDASCPKVNVTCTKHTYRCANGCLSKNGPECDGKEDCGSDGDEK 600  
 DB 541 SQQCNKGDDCGDSDASCPKVNVTCTKHTYRCANGCLSKNGPECDGKEDCGSDGDEK 600  
 QY 601 DCDGGLRSFTRQARVVGTDADBEQWQVSLHALGQGHICGASLISPMLVSLAHAYCID 660  
 DB 601 DCDGGLRSFTRQARVVGTDADBEQWQVSLHALGQGHICGASLISPMLVSLAHAYCID 660  
 QY 661 DRGRYSDPTQMTWFLGLHDQSQTSAPGVQERRLKRIISHPFNDPTFDYDIALLEDEKP 720  
 DB 661 DRGRYSDPTQMTWFLGLHDQSQTSAPGVQERRLKRIISHPFNDPTFDYDIALLEDEKP 720  
 QY 721 AEVSSMRPICLPASHVFPAGKAIWTGWHOTYGGTGAILLKGELRVNNTTCENTL 780  
 DB 721 AEVSSMRPICLPASHVFPAGKAIWTGWHOTYGGTGAILLKGELRVNNTTCENTL 780  
 QY 781 PQQITPRMVCVFLSGVDSCQDGGSGGLSLSSVEADGRIFGAGVYVWGDGCAGRKKPGYTT 840  
 DB 781 PQQITPRMVCVFLSGVDSCQDGGSGGLSLSSVEADGRIFGAGVYVWGDGCAGRKKPGYTT 840  
 QY 841 RLPLFRDMIKENTGV 855  
 DB 841 RLPLFRDMIKENTGV 855

RESULT 2  
 ID ST14 MOUSE STANDARD; PRT; 855 AA.  
 AC P56677;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-JUN-2004 (Rel. 44, Last annotation update)  
 DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Epithlin).  
 GN Name=St14; Synonym=Prss14;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;  
 RX MEDLINE=99216440; PubMed=1019918;  
 RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,  
 RA Schwartz R.H.;  
 RT "Cloning and chromosomal mapping of a gene isolated from thymic  
 RT stromal cells encoding a new mouse type II membrane serine protease,  
 RT epithlin, containing four LDL receptor modules and two CUB domains.";  
 RL Immunogenetics 49:420-428(1999).  
 RN [2]  
 RP REVISIONS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.  
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;  
 RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast tumor;  
 RX MEDLINE=12477932; PubMed=101073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Folley J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shcherenko Y., Bouffard G.G.,  
 RA Blakeley R.M., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalak U., Smallus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -1- TISSUE SPECIFICITY: Highly expressed in intestine, kidney, lung,  
 CC and thymus. Not expressed in skeletal muscle, liver, heart, testis  
 CC and brain.  
 CC -1- SIMILARITY: Belongs to peptidase family S1.  
 CC -1- SIMILARITY: Contains 2 CUB domains.  
 CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL, AF042822; AA02230.3; -;  
 DR EMBL, BC005496; AA05496.1; -;  
 DR HSSP, P00760; 1EZX.  
 DR MEROPS, S01.302; -;  
 DR MGD, MGI:1338881; St14.

DR	GO: 00005576; C:extracellular; IDA.
DR	GO: 0019897; C:extrinsic to plasma membrane; IDA.
DR	GO: 0008236; F:serine-type peptidase activity; IDA.
DR	InterPro: IPR000859; CUB.
DR	InterPro: IPR002172; LDL_receptor A.
DR	InterPro: IPR001254; Peptidase S1.
DR	InterPro: IPR001314; Peptidase S1A.
DR	InterPro: IPR009003; Pept_Ser_Cys.
DR	Pfam: PF00431; CUB; 2.
DR	Pfam: PF00057; Ldl_recept_a; 4.
DR	Pfam: PF00089; Trypsin; 1.
DR	PRINTS: PR00722; CHYMOTRYPSIN.
DR	PRINTS: PR00261; LDLRECEPTOR.
DR	SMART: SM00042; CUB; 2.
DR	SMART: SM00192; Ldl_a; 4.
DR	SMART: SM00020; Tryp_spec; 1.
DR	PROSITE: PS01180; CUB; 2.
DR	PROSITE: PS01209; LDLRA_1; 2.
DR	PROSITE: PS00068; LDLRA_2; 4.
DR	PROSITE: PS50240; TRYPSIN_DOM; 1.
DR	PROSITE: PS50134; TRYPSIN_HIS; 1.
DR	PROSITE: PS00135; TRYPSIN_SER; 1.
KM	Glycoprotein; Hydrolyase; Repeat; Serine protease; Signal-anchor transmembrane.
KW	DOMAIN 1 55
FT	TRANSMEM 56 76
FT	DOMAIN 77 855
FT	DOMAIN 214 331
FT	DOMAIN 340 444
FT	DOMAIN 451 488
FT	DOMAIN 489 522
FT	DOMAIN 523 561
FT	DOMAIN 565 604
FT	DOMAIN 615 854
FT	ACT_SITE 656 656
FT	ACT_SITE 711 711
FT	ACT_SITE 805 805
FT	CARBOHYD 107 107
FT	CARBOHYD 302 302
FT	CARBOHYD 365 365
FT	CARBOHYD 421 421
FT	CARBOHYD 489 489
FT	CARBOHYD 772 772
GO	855 AA; 94654 MW; 4F10E84DA2A146D5 CRC64;

Query Match	82.9%	Score 3884	DB 1	Length 855
Best Local Similarity	81.5%	Pred. No. 4.6e-260		
Matches 697; Conservative	73	Mismatches 85	Indels 0	Gaps 0

Qy	1	MGSDPARCGGGPPR	FCGAGLKTNSRHEKNMGLEBEGYEP	PUNNVKVEKXGPERWVTLA	60	
		1	MGSNNGRAGGGSDP	GAGLKTNSRIENNNGFEEBGEYEP	PANNAKKVEKRGPPRRWVLA	60
Db						
Qy	61	VLIGLLVLLG	LIGLPWHLQYRDV	RVQKXNGMYRTTNEFVDAYENSNSTEFVSLASKY	120	
Db	61	VLFSELTLLIMAGLLV	HEHFNHRYNRVQK	FPNGHLRTTNEFLFDAYENSNSTEFVSLASKY	120	
Qy	121	KDALKLTLSGVPEFL	PRYHKESAVTAFSEBSVLA	AYWSEFSIPQHLVEEAEKRWAAEERYVA	180	
Db	121	KEALKLTLSNEVYLG	RYHKKSATVAFSEBSVLA	AYWSEFSIPRHLAEEDRAAAEVERVAT	180	
Qy	181	LPPRRASLSKSF	VVTSVVAEPFDSKTYQRT	ODNSCSFGLHARGVEIMRFTTTPGPDSPYPA	240	
Db	181	LPPRARALKSF	VLTSVVAEPIDPRMLQRT	ODNSCSFALAHGAAYVTRFTTPGPPNSPYPA	240	
Qy	241	HARCCOMALRGD	ADSVLSLTFERSFEDLAS	CDERSGLVTYVNTLSMPMRHLYOLCGTYPPS	300	
Db	241	HARCCOMALRGD	ADSVLSLTFERSFEDVA	CDERHSDLVTVDSLSMPMRHVAVYLCGTFSPS	300	
Qy	301	YNLTFESSQNV	LLTLINTTERRRHP	FEATFQLPFMSSCGGLRRAQGTENSPPYPGHX	360	
Db	301	YNLTFESSQNV	LLTVLTTWTRRRHP	FEATFQLPFMSSCGGLRSTQGTFFSPYPGHX	360	

QY	36	PNPINDCMNIEVNNNOHNVAREKFFYLLEPGPAGTCKDVJVEINGEYCEERSQFVTS	420
Db	361	PEPINTCTMNIKCPNNNNVAVRKFLEYLDVNPVWSCTKDVJVEINGEKYCEERSQFVSS	4220
QY	421	NSNKITVRPHSDOSYTDITGLAEYLISYDSDFCPGQFTCRTGRCIRKELRCDMADCTDH	480
Db	421	NSSKITVHFPHSDHSYTDITGLAEYLISYDSNDPCPFMFCKTGRCIRKELRCDMADCPDY	480
QY	481	SDELNCSAAGHQFFCCKNKCKPLFMWCDVNDVDCGSDNDBEGSCSPROTFFCSNGKCLSK	540
Db	481	SDEYTCCKNATHQFTCKNQFCRPLFMWCDVNDVDCGSDSBEGSCSPHOSFCSNGKCLPQ	540
QY	541	SOOCNGKDCDGSDESDASCPKVVNVYTTCTKHTYRLCLNGSLCLSKNPECDEKEDCSGSD	600
Db	541	SOCKNGKNDGSDSDEASCDSVNVVYSCTKYTYRCNGSLCLSKNPECDEKTDGSDSDEK	600
QY	601	DCDCGLRSFFRQARVVGCTDADLEGEMPMQVSIHALGCHITGASLISPMNLVSAHAHYID	660
Db	601	NCCCGRSFFRQARVVGCTADLEGEMPMQVSIHALGCHILGASLISPMNLVSAHCFQD	660
QY	661	DROGRYSDDPOMTAFGLGHDQSORBAPGOEERLKRILSHBFFNFETFDYDIALLELEK	720
Db	661	DKMFKISDLYMTWTAFLGLDQSKRBSAVQELKLRILITHHSFNDFPFDDYDIALLELEKS	720
QY	721	AEYSMWVRPICLPDASHVFPAGKAIWVTGMHTQYGGTGALLDQGBELRVYNQTTCENL	780
Db	721	VEYSVTVRPICLPDATHVFPAGKAIWVTGMHTKEGGTGALLDQGBELRVYNQTTCEMDM	780
QY	781	POOITPRMNCVGLSGGVNDCSGDGGSPPLSSVEADGRIFGAGVYVSWBDCAGRNKPEVYT	840
Db	781	POOITPRMNCVGLSGGVNDCSGDGGSPPLSSAEMDGRMFOAGVYVSWBEGCAGRNKPEVYT	840
QY	841	RLPLFRDMIKENTGV	855
Db	841	RLPVRDMIKENTGV	855

```

RESULT 3
BAC35125
ID BAC35125 PRELIMINARY; PRT; 855 AA.
AC BAC35125;
DT 14-APR-2004 (TREMBlrel. 27, Created)
DT 14-APR-2004 (TREMBlrel. 27, last sequence update)
DT 14-APR-2004 (TREMBlrel. 27, last annotation update)
DE 0 day neonate kidney cDNA, RIKEN full-length enriched library,
DE clone: D630041B06 product: suppression of tumorigenicity 14 (colon
DE carcinoma), full insert sequence.
OS
OS Mus musculus (Mouse) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=22354683; PubMed=12466851;
RA THE FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
```

RT "High-efficiency full-length cDNA cloning."  
 RL Mech. Enzymol. 303:19-44(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes."  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitsuana T., Tashiro H., Itoh M.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama N., Nishine T., Harada A.,  
 RA Yamamoto S., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RA Adachi J., Aizawa K., Akimura T., Hara A., Hashizume W.,  
 RA Fukuda S., Furuno M., Hanamura T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,  
 RA Hori F., Imocani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK052738; BAC35125.1; -  
 SQ SEQUENCE 855 AA; 94654 MW; 4F10B84DA2146DD5 CRC64;

Query March 82.9%; Score 3884; DB 2; Length 855;  
 Beet Local Similarity 81.5%; Pred. No. 4.6e-260;  
 Matches 697; Conservative 73; Mismatches 85; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGPPDPFGAGLKYNSRHEKYNGLIEGVEFLPVNNVYKVEKPGRVVLA 60  
 DB 1 MGSNRGRKAGGSGDPFGAGLKYNSRHEKYNGLIEGVEFLPVNNVYKVEKPGRVVLA 60  
 QY 61 VLIGLLVLVLIGLIGLVHLYQYRDVAVQYKQKQYKMTINENFVDAVENSSTPEVSLASKV 120  
 DB 61 VLFFLLSTLMAAGLLVHMFHYRNVAVQYKFNGLRITWEIFLDAVENSTSTPEVSLASQV 120  
 QY 121 KDAKLKLYSGVPFGCPYHKEASAVTAFAFGSVIAYVWSEFSYPOHLYEAEVMAEERVY 180  
 DB 121 KEAKLKLNEVPVIGPYHKSAVTAFAFGSVIAYVWSEFSYIPHLAEVDRAMAERVVT 180  
 QY 181 LPPARSLKSVYVTSVVAFPDTSKTVQRTQNSCSFGIHAAGVLEIMRTGPPDPSYPYA 240  
 DB 181 LPPARSLKSVYVTSVVAFPDPRLOKRTQNSCSFGIHAAGAVTRITGTFPSPYPA 240  
 QY 241 HARCQWALRGADSVLSTFRSPDLASCDEBGSGLVTVYNTLSPMEPHALVOLCGTTPS 300  
 DB 241 HARCQWALRGADSVLSTFRSPDLASCDEBGSGLVTVYNTLSPMEPHAVRLCGTSPS 300  
 QY 301 YNLTFHSSQNTLLTLINTBRHRCFEATFPOLPRMSCCGRRLKAGCTNSPYVPGHY 360  
 DB 301 YNLTFHSSQNTLLTLINTBRHRCFEATFPOLPRMSCCGRFLSDTGITSSPYVPGHY 360  
 QY 361 PPNIDCTNNIEVPPNNQHVRFKFFVYLLPGVPACTCPKDYVEINGEKYCGERSQFVTS 420

DB 361 PPNINCTNWKIKVPNNRAVYKRFKLYLVDPNVVPGSCTKDYVEINGEKYCGERSQFVSS 420  
 QY 421 NSNKITVRFPSDOSYDTGTGLAEYLSYDSSDPGCGTCTGRCIRRELBQDGAADCTD 480  
 DB 421 NSSKITVRFPSDHSYDTGTGLAEYLSYDSSNDPCGMFMCKTGRCIRRELBQDGAADCPDY 480  
 QY 481 SDEINSCSDAGHOFTCKNFKCKPLFWYCDSDVNDGSDSDGSCSPAQTRFCSNGKCLSK 540  
 DB 481 SDEHYCCNMTHTQTCNQCKPLFWYCDSDVNDGSDSDGSCSPAGSRKCSNGKCLPQ 540  
 QY 541 SQQCNHGDCGSDGSDASCPKVVVVTCTKTYRCLNGLCLSKNGPECDGSDGSDG 600  
 DB 541 SQKCNHGDCGSDGSDASCPVVVVSCTKTYRCQNGCLCLSKNGPECDGSDGSDG 600  
 QY 601 DCDGGLASFTROARVVGSTADDEGEWQVSLHALGGHICGASLISPNMLVSAHICYD 660  
 DB 601 NCDGGLASFTROARVVGSTADDEGEWQVSLHALGGHICGASLISPMVLVSAHICFQD 660  
 QY 661 DRGFRSDPTQWTFPLGLHPQSORAPGVORLKRITISHPFNDFFDYDIALLELEXP 720  
 DB 661 DKMFKYDITWMTAFELGLDQSKRSAGVDELKRIITHPSFNDFFDYDIALLELEKS 720  
 QY 721 AEYSSMVRPCLPDASHVPAGKAIWYTGNGHTQYGGTGAIIQKGEIRVYNQTCENLL 780  
 DB 721 VEYSTVVRPCLPDATHVPAGKAIWYTGNGHTGEGGTGAIIQKGEIRVYNQTCEDLM 780  
 QY 781 PQQITPRMVCVGLSGVDSGQDGGPLSSVEADGIFPAGVYWMGDGACGRKPGVYT 840  
 DB 781 PQQITPRMVCVGLSGVDSGQDGGPLSSVEADGIFPAGVYWMGDGACGRKPGVYT 840  
 QY 841 RLPLFRDMIKENTGV 855  
 DB 841 RLPLFRDMIKENTGV 855

RESULT 4  
 ID 09JUI7 PRELIMINARY; PRT; 855 AA.  
 AC 09JUI7  
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)  
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)  
 DT 05-JUN-2004 (T-EMBLrel. 27, Last annotation update)  
 DE Membrane bound serine protease (Membrane bound arginine specific  
 DE serine protease).  
 GN Name=MBSP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=wislar; TISSUE=Jejunum;  
 RX MEDLINE=21458307; PubMed=11573963;  
 RA Sacomi S., Yamagaki Y., Tanuzaki S., Hitomi Y., Iwanaga T., Fushiki T.;  
 RT "A role for membrane-type serine protease (MT-Sp1) in intestinal  
 RT epithelial turnover".  
 RL Biochem. Biophys. Res. Commun. 287:995-1002(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=wislar; TISSUE=Ductum;  
 RX Inoue H., Takahashi K., Kishi K.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC 1. SIMILARITY: Belongs to peptidase family 51.  
 DR EMBL; AB037898; BAB03502.1; -  
 DR PIR; JCT731; JCT731.  
 DR HSSP; P00760; IEZX.  
 DR MEROPS; S01.302; -  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000859; CUB.

DR InterPro: IPR002172; LDL\_receptor\_A.  
 DR InterPro: IPR001254; peptidase\_S1.  
 DR InterPro: IPR001314; peptidase\_S1A.  
 DR InterPro: IPR009003; Pept\_Ser\_Cys.  
 DR Pfam: PF00431; CUB; 2.  
 DR Pfam: PF00057; Ldl\_recept\_a; 4.  
 DR Pfam: PF00089; Trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PRINTS: PR00261; LDLRECEPTOR.  
 DR PROSITE: PS01180; CUB; 2.  
 DR PROSITE: PS01209; LDLRA\_1; 2.  
 DR PROSITE: PS50068; LDLRA\_2; 4.  
 DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 855 AA; 94955 MW; 35806B7E6CF6CF03D CRC64;

Query Match 82.6%; Score 3866; DB 2; Length 855;  
 Best Local Similarity 80.8%; Pred. No. 8e-259;  
 Matches 691; Conservative 79; Mismatches 85; Indels 0; Gaps 0;

QY 1 MGSDBAKGGGGPRDPGAGLKYNSRHRKNGLEGEVFLPVNNKTKYKXGPGRWVYLA 60  
 DB 1 MGNNGRKGAGGGSDPFAGLKYNRLNMGFEHGVFLPVNNAKQYKXGPRRWVYVA 60  
 QY VLGILGLVLLGIFLVHMLQYRDYRQKYNKGRINENFVDAEYNSNSTEYFVSLASKV 120  
 DB VVPSFLLSLMAGLLVHMFHYRNRVRIQKVRNGHRLINENFLDAEYENSTSTETPSLASQV 120  
 QY 121 KDALKLYSGVPLGPRYHKESAVTAPEBGSYIATYWGSEFSIPOHLVEAEERWMAEERVVM 180  
 DB 121 KEALKMYSEVPVIGPRYHKSTVAFSEGSYIATYWGSEFSIPHLVEEVRAMAEERVVT 180  
 QY 181 LPPRABSLKSFVYTVSVAFPTDSKTQRTQDSCSFLHARGVLMFTTGPSPDSYPA 240  
 DB 181 LPPRABSLKSFVYTVSVAFPTDPMQRTQDSCSFLHARGVLMFTTGPSPDSYPA 240  
 QY 241 HARCQMLARGDAUSVLSLTFERSFDLASCDESGSDLVVYNTLSPEMHPALVOLCGTYPSP 300  
 DB 241 HARCQWLARGDAUSVLSLTFERSFDLASCDESGSDLVVYNTLSPEMHPALVOLCGTYPSP 300  
 QY 301 YNLTFFSSQNVLLTTLTTERRHPRGEATFPOLPRMSGCGGRILKAQGFNPSYYRGHY 360  
 DB 301 YNLTFFSSQNVVLTTLTTERRHPRGEATFPOLPRMSGCGGRILKAQGFNPSYYRGHY 360  
 QY 361 PPNIDCTWNIENVNNOVKVRFKFFVLEEGVPRGCPKQYVEINSGKYGESGQFVYTS 420  
 DB 361 PPNIDCTWNIENVNNOVKVRFKFFVLEEGVPRGCPKQYVEINSGKYGESGQFVYTS 420  
 QY 421 NSNKTIVRFHSDSYTDYTGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCDGWADCTDH 480  
 DB 421 NSNKTIVRFHSDSYTDYTGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCDGWADCTDH 480  
 QY 481 SDELINSCDAGHQTCKNKKFCKPLFWYCDVYNDGDSDEGSCCPAOTRCSNGKLSK 540  
 DB 481 SDEHHCRCNATHQMCCKQFCKPLFWYCDVYNDGDSDEGSCCPAOTRCSNGKLSK 540  
 QY 541 SQQNGKDDGDSGDSASCPKVVNVYCTKHTYRCLNGLCLSKNGPECDGKEDSDGSD 600  
 DB 541 SQQNGKDDGDSGDSASCPKVVNVYCTKHTYRCLNGLCLSKNGPECDGKEDSDGSD 600  
 QY 601 DCDGSLRSTROARVVGTDADGEMPMOVSILHAGGHI CGASLISPNMLVSAHACYID 660  
 DB 601 NCCGSLRSTROARVVGTDADGEMPMOVSILHAGGHI CGASLISPNMLVSAHACYID 660  
 QY 661 DRGRISDPLQWTAFLGLHDSQSAAGVQERRILKRIISHPFNDFFPDIALLELEKP 720  
 DB 661 ETTFKXSDHMTWTFALGLDQSKASAGVQENKIKRIITPSPFNDFFPDIALLELEKP 720  
 QY 721 AEYSAMWRPCLPRASVVPAGKAIWYTGNGHIOYGGTGLIIQXGIRVYNQTTCNL 780  
 DB 721 AEYSVVRPCLPRNTHVFPAGKAIWYTGNGHIOYGGTGLIIQXGIRVYNQTTCELL 780

QY 781 PQQITPRMVCVGLSGGVDSQCQDSCGSPUSVEADGRIFGAVGVSWGDCGAGNKGVYT 840  
 DB 781 PQQITPRMVCVGLSGGVDSQCQDSCGSPUSVEADGRIFGAVGVSWGDCGAGNKGVYT 840

QY 841 RLPLFRDMTKENTGV 855  
 DB 841 RIPEVRDMKEQRTGV 855

RESULT 5  
 ID Q6GR54 PRELIMINARY; PRT; 845 AA.  
 AC Q6GR54;  
 DT 05-JUL-2004 (TEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TEMBLrel. 27, last sequence update)  
 DT 05-JUL-2004 (TEMBLrel. 27, last annotation update)  
 DE Scl4-A-prov protein.  
 GN Name=scl4-A-prov;  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;  
 OC Xenoportidae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Stennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marziska K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guinardne P.H.,  
 RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,  
 RA Kzyvinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Maxia M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22341132; PubMed=12454917;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.;  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT Initiative.";  
 RL Dev. Dyn. 225:384-391 (2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.;  
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
 CC - SIMILARITY: Belongs to peptidase family S1.  
 DR EMBL: BC071077; AH71077.1; -;  
 DR InterPro: IPR000859; CUB.  
 DR InterPro: IPR002172; LDL\_receptor\_A.  
 DR InterPro: IPR001254; peptidase\_S1.  
 DR InterPro: IPR001314; peptidase\_S1A.  
 DR InterPro: IPR009003; Pept\_Ser\_Cys.  
 DR Pfam: PF00431; CUB; 2.  
 DR Pfam: PF00057; Ldl\_recept\_a; 4.  
 DR Pfam: PF00089; Trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.

DR PRINTS; PRO0261; LDRRECEPTOR.  
DR SMART; SM00042; CUB; 2.  
DR SMART; SM00192; LDLa; 4.  
DR SMART; SM00020; TRY\_SPC; 1.  
DR PROSITE; PS01180; CUB; 2.  
DR PROSITE; PS01209; LDLa\_1; 2.  
DR PROSITE; PS50068; LDLa\_2; 4.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR HydroLase; Protease; Serine protease.  
SQ SEQUENCE 845 AA; 93593 MW; B917386C95BC73BD CRC64;

Query Match 57.4%; Score 2686; DB 2; Length 845;  
Best Local Similarity 55.7%; Pred. No. 3.7e-177; Matches 472; Conservative 154; Mismatches 212; Indels 10; Gaps 5;

QY 14 KDFGAGLKYSRHRKVNGLSEGEVEFLPVNNVKKYKHKPGRWVLAVALIGLLVLLGIG 73  
DB 2 KDSMMKYNRRPQSLNGFEGVEFLPATNSKYEKTPKKKLAIFGLVIGALLSLTIG 61  
QY 74 FLVWHLQYRDVAVOKVNGKRTINENFVDAIYENSSTEFVSLASKYKDALKLYSGVVF 133  
DB 62 LTVWHPFVYRNKPVNKLTYGYLTIANTPFIDAYENSTTAEFSDLSAKVIDTLQTYVNGKND 121  
QY 134 LGRPHKESATVAFSEG---SVIAYWSEPSIQHLVEBEAEVMAEERVWMLPPRAISLKS 190  
DB 122 IAPLQKCSISAFSEGGNNVIGYWSEFVPAREAAFEAISLKLPSVNPOR--T 178  
QY 191 FVTVSVVAFPTDSKTVDRTQDNCSCFGLHARGVELMFTTPGFPDSPYPAHARCOMALRG 250  
DB 179 FALDSLVAFPTDPIAVFKNSSCAVFLHSSNGVAVAFSSGFPDSEPRARACLMTLRA 238  
QY 251 DADSVLSLTFERSPLASCDERGSOLVTVYNTLSMEPHALVOLCGTYPPSTNTLTFSSQN 310  
DB 239 DAGRIHLHFTFKMECKKPMGSGFVWVYDLSIEPPAQIRLGLIYPSNTLTFSSSN 298  
QY 311 VLLITLTNTERHRHGFEEATFEOFLPRMSSCGRLRKQGTNSPYRPHYPPNIDCTWNI 370  
DB 299 VMLVTLVTDVNGKPPGFLAEFKQLPKTSLCCGLRDSAGFTSPFPAHPSTECIMDI 358  
QY 371 EVPNNAQVKKVAFKFFYLEPGVAPAGTCPKDYVEINGEKYCGERSQFVTVSNKITYRFH 430  
DB 359 QVPENKVKKVAFNNFYLAPEGVPTKCTKDPVEIKGQCYCEKFPFVSNNSKMSGRFV 418  
QY 431 SDQSYDTGFLAETLSDSDPCGQFTCRGCIKRELKCDGADCTDHSDELNCSDA 490  
DB 419 SDQSYDTGFLAETLSDSDPCGQFTCRGCIKRELKCDGADCTDHSDELNCSDA 478  
QY 491 GHOTCKK-KFCKRLFWVCSVNDGCGNSDQSGSCPAQTFRCNGKCLSKSQCNKDD 549  
DB 479 -LQRCVNSKLCCKSFYPCDGVNCGSSDDELACKCPNNTPFKCNGKCIPTDQKCDRVN 537  
QY 550 CGDSDSDASCPKVVVCTKATYRCLNGLCISKNPEDCKEDSCDSDEK--CCDCGLR 607  
DB 538 CGDSDSDASCPKVVVCTKATYRCLNGLCISKNPEDCKEDSCDSDEK--CCDCGLR 597  
QY 608 SFTROARVVGCTDADEGEWVQVSLHAGQGHICGASLISFNMVLSAAHCYIDRGRFYS 667  
DB 598 PFTKKSRIYGVGNADTEGFPMQVSLHAKNGHGTGASIVSPTMLISAHCQDHDHQRYS 657  
QY 668 DPQWMTFLGLHDSQSRASGVORERLKRITISHPFNDFPDVIALLEKPAEYSMV 727  
DB 658 DASIMTYLGLHDDAQINTKDVVERKIKRIMAHIGFNDNTYDNDIAVLELEKPEVYDFPI 717  
QY 728 RPICLPASHVFPAGKAIWVTGMGHTQYGTGALLQKGEIRVINGTTCENTLPOQITPR 787  
DB 718 QPVCIPSTHDFPVGKPIWVTGMGALKKEGGAIVILQKAEIRIINGTECNLDGQILPR 777  
QY 788 MCVGFLSGGVSDCGDSDGGLSVSEVADGRIFGAGVSVSGDCGGRNKPQYITLPLFRD 847  
DB 778 MLCAGVSGGIDACCGDSDGGLSVSEVLENNKVVYLAGVSVSGDCGGRNKPQYITLPLFRD 837

QY 848 WIKENTGV 855  
DB 838 WIKDXTGL 845

RESULT 6  
ID Q9DGR1 PRELIMINARY; PRT; 845 AA.  
AC Q9DGR1; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Homolog of human MT-SPI.  
GN Name=MT-SPI;  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20363741; PubMed=10903452;  
RA Yamada K., Takabatake T., Takeshima K.;  
RT "Isolation and characterization of three novel serine protease genes  
from Xenopus laevis."  
RL Gene 252:209-216 (2000).  
CC -1- SIMILARITY: Belongs to peptidase family S1.  
DR EMBL; AB038498; BAB08218.1; -.  
DR HSSP; P00760; 3BTH.  
DR MEROPS; S01.050; -.  
DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
DR GO; GO:0004295; F:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR002172; LDU\_receptor\_A.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR009003; pept\_ser\_Cys.  
DR Pfam; PF00431; LDU\_2.  
DR Pfam; PF00057; LDU\_recept\_a; 4.  
DR Pfam; PF00089; Trypsin\_1.  
DR PRINTS; PRO0261; CHYMOTRYPSIN.  
DR PRINTS; PRO0261; LDRRECEPTOR.  
DR SMART; SM00042; CUB; 2.  
DR SMART; SM00192; LDLa; 4.  
DR SMART; SM00020; TRY\_SPC; 1.  
DR PROSITE; PS01180; CUB; 2.  
DR PROSITE; PS01209; LDLa\_1; 2.  
DR PROSITE; PS50068; LDLa\_2; 4.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR HydroLase; Protease; Serine protease.  
SQ SEQUENCE 845 AA; 93597 MW; 7FD7E62851A758B7 CRC64;

Query Match 56.8%; Score 2660; DB 2; Length 845;  
Best Local Similarity 55.4%; Pred. No. 2.4e-175; Matches 470; Conservative 154; Mismatches 216; Indels 10; Gaps 5;

QY 14 KDFGAGLKYSRHRKVNGLSEGEVEFLPVNNVKKYKHKPGRWVLAVALIGLLVLLGIG 73  
DB 2 KDSMMKYNRRPQSLNGFEGVEFLPATNSKYEKTPKKKLAIFGLVIGALLSLTIG 61  
QY 74 FLVWHLQYRDVAVOKVNGKRTINENFVDAIYENSSTEFVSLASKYKDALKLYSGVVF 133  
DB 62 LTVWHPFVYRNKPVNKLTYGYLTIANTPFIDAYENSTTAEFSDLSAKVIDTLQTYVNGKND 121  
QY 134 LGRPHKESATVAFSEG---SVIAYWSEPSIQHLVEBEAEVMAEERVWMLPPRAISLKS 190  
DB 122 IAPLQKCSISAFSEGGNNVIGYWSEFVPAREAAFEAISLKLPSVNPOR--T 178  
QY 191 FVTVSVVAFPTDSKTVDRTQDNCSCFGLHARGVELMFTTPGFPDSPYPAHARCOMALRG 250



```

Db 179 FALDSLVAEPTDPOIAFVFNKNSCAVFLHSSNGVAFSSPGPSPDPYPPNARCLWTIRA 238
Qy 251 DADSVSLFTRSPFLASCDERGSGLVYVNTLSMPMEHALVOLCGTTPSPYNTLFHSSON 310
Db 239 DAGRIIHLHRTKFMCKCPNGGDFFVNVYDLSLTERPAQRLCGIYPPSTLTFSSSN 238
Qy 311 VLLITLITNTERRRHGPFEATFPOLRPMSSCGRLKKAQGTFSNPPYRGHYPNIDCTWNI 370
Db 299 VMLVTLVTDVNVGKPGFLAEFKQLPKTSLCGGILRDASGFTSPYPAHYPPSTESIMDI 358
Qy 371 EVPRNNGVAKRPFKFFYLLEBGVAGTGPXDYVEINSGKYGSGSQPVYTSNSKITYREFH 430
Db 359 QVBDNKKFVKRPFMVFYLAEBGVVPYTKCTDVEIKGQYGEKGFVVSNNSSKMSYRFV 418
Qy 431 SDQSYDTGTFLAEVLSVSDSDPCGQGTCTGRCIRKELSCDGDADCTDSDSLNCSGDA 490
Db 419 SDQSYDTGTFLAEVLSVSDSDPCGQGTCTGRCIRKELSCDGDADCTDSDSLNCSGDA 478
Qy 491 GHQFTCKN-KECKPLFWVCDSDVNDGDNSSDEGSCSPAQTFRGSCGKLSKQOQNGKDD 549
Db 479 -LQPRCVNSHLCKRSPYFCGVDVNDGDNSSDEGSCSPAQTFRGSCGKLSKQOQNGKDD 537
Qy 550 CGGSGSDASCPKRVNVVTCTHTYRCLNGLCLSKNPECDKEDCSGSDSK--DCDQGLR 607
Db 538 CGGSGSDAECDQVLYTCTEYTKCKNNQCTYKKNPCDENDSCSDSDENAAKCKNGKR 597
Qy 608 SPTROARVVGCTDADGEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDRGFRYS 667
Db 598 PFTKKSIVGVANADTEPFPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDRGFRYS 657
Qy 668 DPTQWTAFLGIDHQSASABQVQERLKRISHPFNDFTFDYDIALLEKPAEYSSMV 727
Db 658 DASIMTAYLGLHDOAQNTXDQVVERIKRIMAHIGFNDNTYDNDIAVLEKPEYETDPI 717
Qy 728 RPLCTPLASIVFPAGKAIWMTGHTGYGSGALILQKEIRVYNQTTGENTLPOQITPR 787
Db 718 QPVCIPPESTHDFVKGKPIWGTGALKEGGAAVILQKAIRILINQTECKNLDGQITPR 777
Qy 788 MMCVGFSLSGVSDSCGSGGPGPLSSVEADGRITFGAVVSMWGDGAGRRKPGVYTRLPFRD 847
Db 778 MLCAGFVSGSIDACQSGSGPLSSVELNNKYLAGVSMWGDGAGRRKPGVYTRKSMRD 837
Qy 848 WIKENTGV 855
Db 838 WSKDKITGL 845

RESULT 7
Q8WVC1 PRELIMINARY; PRT; 422 AA.
ID Q8WVC1 AC Q8WVC1;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE S114 protein (Fragment).
GN Name=S114;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=2388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusik K., Farmer A.A., Rubin A.G., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

```

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,
RA Villalon D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smallin D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maria M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
CC -!- SIMILARITY: Belongs to the EMBL/Genbank/DBJ databases.
DR EMBL; BC018146; AAH18146.1; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00089; Trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 4.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS50068; LDLRA_2; 4.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON TER
SQ
SEQUENCE 422 AA; 46257 MW; 2C99875D1B58B319 CRC64;

Query Match 50.5%; Score 2365; DB 2; Length 422;
Best Local Similarity 99.5%; Freq. No. 2, 6e-155;
Matches 420; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 434 SYTDGFLAEVLSVSDSDPCGQGTCTGRCIRKELSCDGDADCTDSDSLNCSGDAHQ 493
Db 1 SYTDGFLAEVLSVSDSDPCGQGTCTGRCIRKELSCDGDADCTDSDSLNCSGDAHQ 60
Qy 494 FTCKNFKCPLEFVWCDSDVNDGDNSSDEGSCSPAQTFRGSCGKLSKQOQNGKDDCG 553
Db 61 FTCKNFKCPLEFVWCDSDVNDGDNSSDEGSCSPAQTFRGSCGKLSKQOQNGKDDCG 120
Qy 554 SDEASCPKRVNVVTCTHTYRCLNGLCLSKNPECDKEDCSGSDSKDDCCGSRFTROA 613
Db 121 SDEASCPKRVNVVTCTHTYRCLNGLCLSKNPECDKEDCSGSDSKDDCCGSRFTROA 180
Qy 614 RVVGGTDADEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDRGFRYSDEPTQW 673
Db 181 RVVGGTDADEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDRGFRYSDEPTQW 240
Qy 674 AFLGLHDQSORSPAVQOERLKRISHPFNDFTFDYDIALLEKPAEYSSWVRDCLP 733
Db 241 AFLGLHDQSORSPAVQOERLKRISHPFNDFTFDYDIALLEKPAEYSSWVRDCLP 300
Qy 734 DASHVPPAGKAIWMTGHTGYGSGALILQKEIRVYNQTTGENTLPOQITPRMWCVP 793
Db 301 DASHVPPAGKAIWMTGHTGYGSGALILQKEIRVYNQTTGENTLPOQITPRMWCVP 360
Qy 794 LSGGVNSCGSDSGPPLSSVEADGRITFGAVVSMWGDGAGRRKPGVYTRLPFRDWIKENT 853
Db 361 LSGGVNSCGSDSGPPLSSVEADGRITFGAVVSMWGDGAGRRKPGVYTRLPFRDWIKENT 420
Qy 854 GV 855

```



Db 421 CV 422

RESULT 8

Q6DEVO PRELIMINARY; PRT; 663 AA.

AC Q6DEVO; ID Q6DEVO; DT 01-OCT-2004 (TREMblrel. 28, Created); DT 01-OCT-2004 (TREMblrel. 28, Last sequence update); DT 01-OCT-2004 (TREMblrel. 28, Last annotation update); DE Hypothetical protein.; OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis); OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8364;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RX MEDLINE=22388257; PubMed=12477932;

RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedin T.B., Toshiyuki S., Carrinci P., Prange C., Rana S.S., Loguellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiteing M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez R.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywnski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E., Jones S.J., Maitra M.A., "Generation and initial analysis of more than 15,000 full-length human RT and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RA Klein S., Strauberg R.; Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.

RL EMBL; BC076994; AAH76994.1; -.

KM Hypothetical protein.

SQ SEQUENCE 663 AA; 73914 MW; 46B2A56C657C7739 CRC64;

Query Match 42.0%; Score 1965; DB 2; Length 663;

Best Local Similarity 42.9%; Pred. No. 2.2e-127;

Matches 364; Conservative 120; Mismatches 172; Indels 192; Gaps 6;

QY 14 KDPGAGLKYNGRHRKVNLEEGVEFLPVNNKYKYEKGKPGWVLAANTLIGLLVILGIG 73

DB 2 KDSUSMMKYNNRPSQMGEEGVEFLPVANNKYKYEKAPKKGLAIFGVVIAALLSLTIG 61

QY 74 FLVWHLQYRDVAVKNGVYRITNENFVDAYENSSTEFVSLASKVDAKLTVSGVPF 133

DB 62 LLVWHLFAVRNAPVQKLTGYIRIANTQVFEAYENSTTEFPDLISKVISTRTIYNGKD 121

QY 134 LGPYHKESAVTAFFEGS--VIAYYWSFSLPQHLVEAEERVMAEERVVMLPPRARSIKS 190

DB 122 IAPYLQCSISAFSEGSDNNVGYWSEFVPAFEAEFEKAISE--LKLPTVNLQRA 178

QY 191 FVTVSVVAFPTDQKTVQTONSCSPGLHARGVLMRTTGTPSPVPAHARQWMLRG 250

DB 179 FAVSVLVAFPTDQKTVQTONSCSPGLHARGVLMRTTGTPSPVPAHARQWMLRG 238

QY 251 DADVSLTTFPSFDLASCDERGSDLVTVYNTLSLPMEPHALVOLCTGYTPSPVNTLTFHSSON 310

DB 239 DAGGMIRLAKFTFKMEKCKANAGDFVWVYDLSLPIEPAAQRLGCIYTPSNLTTFPSSSN 298

QY 311 VILITLITNTERBRHGPGEATFPOLPRMSSCGGRLRKAQGTENSPYPYRGHYPPNIDCTMNI 370

DB 299 VMLTTLVTDNVCKRPGGLAFSPGPKTSLCGYIYRDASGVFTSPDFGHIPTKICTWDI 358

QY 371 EVPNQHVAVRPFKFFYLLBPGVAGTCPKDYVEINGEKYCGERSQFVYVTSNKNITVRFH 430

DB 359 QVPDNKFKVLRFMNFYLAEPGVPTKCTKDVEILINGYKGERKFFVSNSSKMSRVFV 418

QY 431 SDQSYDTGTFLAELYSIDSSDPGQFTCTGRCIRKELACDGMADCTDSDELACSCDA 490

DB 419 SDQSYDTGTFLAELYSIEPPNCPDQFACKSGCIRLDQCDWMNDCEDSDERSCTCTA 478

QY 491 GHOFTCKN-KFCKPLFVVCSDVNDGDNNSDEGSCSPAQFRCNSGKCLSKSQQCNKGD 549

DB 479 -LQPRCTNSKLCRPSYVCDGVNDGSDLSLACQCPNNITKYCGNGKCLPESQKCDRTDN 537

QY 550 CGDSDASCPKVVVYCTGHTYRCLNGLCISKGNPECDKEDCSDSDSDE--KDCDGLR 607

DB 538 CGDSDASCPKVVVYCTGHTYRCLNGLCISKGNPECDKEDCSDSDSDE--KDCDGLR 597

QY 608 SFTROARVVGCTDADEGEWPMQVSLHALGGGHI CGASLISPNMLVSAHICYIDRGRRYS 667

DB 598 PFTKSRIVGVANA----- 611

QY 668 DPTQWTFGLGHDSQSRASRGVQERLRKRIISHPFNFDPFDYDIALLEKPAEYSSWV 727

DB 612 ----- 611

QY 728 RPICLPDASHVFPAGKAIWVTGNGHTQYGGTALLQKGBIRVYNQTTCEMLLPQJTPR 787

DB 612 ----- 611

QY 788 MCVGFLSGVDSQGDSPGLSSVEADGRIFGAGVSWDGCAGNRKPGVYTRLPFRD 847

DB 612 -----DSGGPLSSVEILNRKYLAGIVSWGEGCARRKPQVYTRVAMMRD 655

QY 848 WIKENTGV 855

DB 656 WIKENTGV 663

RESULT 9

Q6PP94 PRELIMINARY; PRT; 799 AA.

AC Q6PP94; ID Q6PP94; DT 05-JUL-2004 (TREMblrel. 27, Created); DT 05-JUL-2004 (TREMblrel. 27, Last sequence update); DT 05-JUL-2004 (TREMblrel. 27, Last annotation update); DE Tmpres6 protein.; OS Mus musculus (Mouse); OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=FVB/N; TISSUE=liver;

RX MEDLINE=22388257; PubMed=12477932;

RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedin T.B., Toshiyuki S., Carrinci P., Prange C., Rana S.S., Loguellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiteing M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Liver;  
 RA Strausberg R.;  
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to peptidase family 11.  
 DR EMBL; BC057674; AAH57674.1; -  
 DR GO; GO:0005886; C:plasma membrane; IDA.  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR002172; IDL\_receptor\_A.  
 DR InterPro; IPR001254; Peptidase\_S1-  
 DR InterPro; IPR003003; Pept\_Ser\_Cys.  
 DR Pfam; PF00089; Trypsin; 1.  
 DR Pfam; PF00057; Ldl\_recept\_a; 2.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR SMART; SM00192; IDLA; 3.  
 DR SMART; SM00202; TRYP\_SPE; 1.  
 DR PROSITE; PS01180; CUB; 1.  
 DR PROSITE; PS01209; LDLRA\_1; 1.  
 DR PROSITE; PS00068; LDLRA\_2; 3.  
 DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 KM Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 799 AA; 89557 MW; 16315A646A4D5288 CRC64;

Query Match 23.8%; Score 114.5; DB 2; Length 799;  
 Best Local Similarity 31.6%; Pred. No. 2,1e-68;  
 Matches 266; Conservative 140; Mismatches 347; Indels 90; Gaps 25;

QY 33 EEVVE---FLPVNVKVKKEKGGKGRVWVLAVALIGLLVLLGIGPLVMH-LQYR-DVRV 86  
 DB 20 EEAEPBGKFKFPRKNTKR-----KNRDVVRTPPLLVLAALVSAAGVLMWYLGKAEVTV 74

QY 87 QKVNKGKMRITNEFVPAVAKNSNTEFVSLASKYKDKALKLYSVPLGPRPKESANTAF 146  
 DB 75 SOVYSGSLKRLVLRNRFSSQDLGRREIARFSEAKAKQKMLQELVASTR-LGYTVNSSVYSF 133

QY 147 SEGSVIAYVSESEFIPQHLVEAEARVAAEERVV--MLPPRARSLKSPVTVSVVAFPTDSK 204  
 DB 134 GEGPLTGFPMFLIDPEY-----ORLTSPVAVBELVDLLSNSTLASKTTEYVDPE 188

QY 205 --TVQRTQDNSCSFGLHARGVLEMRFTTPG--FP-DSPYPAHARCWALRGDADSVLSLT 259  
 DB 189 GLVILEASVNDIVLINSTLGCYRYSVNPQGVLPKGPDDQTTCLMHLQGPEDLMIKVR 248

QY 260 FRSDLASCDERGGDLYTVNYTSLSPMEPHALVQCGTTPSPSYNTLTHSSQNVLLITLTN 319  
 DB 249 L-ETRTVDCRDR---VAAMDAAQPLEKRLTLTVYSGSRQRPVVEVLASGSVAWVWKKG 303

QY 320 TERHPPGF---EATFPQLPRMS--SCGGRLLKAGCTFNSPYYPHYPPNIDCTNIEVFN 374  
 DB 304 MHSYIDFLLSVKSAFQDCQVNLTEGRLL-DTQGLPRLTPYSPYSPSTHCSNHLTVPS 362

QY 375 NQHVKVFKEFFYLEPVPAGTCKPDYVEINGEKCYGER-----SQFVTVSNSKKTIVRF 429  
 DB 363 LDVGLALMFDAVYALRQKYNRLCTQGGWMIQNRLLCGFRTLQPAERLIPMAVASGVITLNF 422

QY 430 HSDQSYDTDTGLAAYLSYDSDPFGQFTCTGTGCIKEKLRCQDMACTDHSDELNSCD 489  
 DB 423 TSQISLTGPGVQVYVSLYNOSDPCGGEFLCV----- 454

QY 490 AGHPTCKNKKCKFLFWVCDVNDGNSDEGSCCPAQTPRC-SNGCKLSKSQCGNGKD 548  
 DB 455 -----NGLCVF---ACDGIKDCPNGLDERNCVCRAM-FQCGEDSTCTISLPRVCDROP 502

QY 549 DCGDSDDEASCPRVNVVTCTKHTYRCLNGLCLSKNGPECDGKEDSGSDGKDCGLRS 608  
 DB 503 DCLNSDEBQCQR--GVPCGTFTFGCEDRSQVKKPPECQGDGDCDSDGDEHCGCLOG 560

QY 609 FTQDARVVGTTDADBEEMPNQVSLHALGCGHTGASLISNMVLSAHCYTDIDRGFRYS 668  
 DB 561 LS--SRVGGTVSBEEMPNQASLQIRGR-HICGGLADRWVITAHACQED--SMAS 614

QY 669 PTQMTAFPLGHDSORSAPGVQERRLKRIISHPPFNDFTFDYDIALLEKPAEVSMMVR 728  
 DB 615 PKMTVPLFGMRQNSR-WPEGVSFYSRLFLPHYHEDSHDYVALLQDHPVYSATVR 673

QY 729 PCLPDAHVFPAGKAIWVTGHTGYGTTGALLQKGLRYVNTQTCENLPLPQQTTPRM 788  
 DB 674 PVCLPARSHFFEEPGCHWTGMCAGOREGFSVNTLQKVQVLPDLCSEAYRYQVSPRM 733

QY 789 MCYGFSLSGVDSQCGSGGSLVSEADGRFFGSGVSWGDCGARRKPGYTRPLPFRDW 848  
 DB 734 LCAGYRKGRKXDCQSDSGGLVCRPPSGRWFLGLVSGWGLGCRPNPFVGYTRVTVYIM 793

QY 849 IKE 851  
 DB 794 IQQ 796

RESULT 10  
 AAH57674 PRELIMINARY; PRT; 799 AA.  
 ID AAH57674  
 AC AAH57674  
 DT 02-MAR-2004 (TREMblrel, 27, Created)  
 DT 02-MAR-2004 (TREMblrel, 27, Last sequence update)  
 DT 02-MAR-2004 (TREMblrel, 27, Last annotation update)  
 DE Trpms6 protein (Mouse).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Liver;  
 RX MEDLINE=22386257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins B.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marsalis K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.U., Abramson R.D., Muliani S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Pabey J., Helton B., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Liver;  
 RA Strausberg R.;  
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC057674; AAH57674.1; -  
 SQ SEQUENCE 799 AA; 89557 MW; 16315A646A4D5288 CRC64;

Query Match 23.8%; Score 114.5; DB 2; Length 799;  
 Best Local Similarity 31.6%; Pred. No. 2,1e-68;  
 Matches 266; Conservative 140; Mismatches 347; Indels 90; Gaps 25;

33 BEGVE-----FLPVANNVKVEKGGPGRWVLAAILGLLVLGIGLVWH-LQYR-DVRY 86  
 20 EEAAPBEKKEPKPKTKR-----KRDYVRFTEPLILVALVSAQWIMWYLGKAEYTV 74  
 87 QKKNQYKRTINENFVDAENSSTEFVSLASKYKDALKLYSGVFLGPKHESAVTAF 146  
 75 SQVYSGSLRVNRFPSQDLGRRESIAFRSESAAKQKMLQELVASTR-LGYTSSSVYSF 133  
 147 SEGVIAYVYSEFSIPQHLVEAEERVAEERVV--MLPPRRLSKSFVYTVAVAPFDISK 204  
 134 GEGPLTGFEPFLIPILPEY-----ORLTISPVEVRELLVDELLSNSTLASVXTEEVDP 188  
 205 --TVQRTQDNCSGFLHARGVELMRFTTSG--FP--DSPYPAHARQWALRGDADSVLSLT 259  
 189 GLVILASVNDIVVNLSTLGCYRSYVNPQVLLPKGDQDTTCLMHLQEPEDIMTVR 248  
 260 FRSPDLASCDERGDVLYVYNTLSPMEHALVOLCGTVPSPYNTLTFHSSQVLLITLTN 319  
 249 L-EWTRVDCRDR-----VAMYDAAGLEKRLITSYVGGSRQBPMEVLASGSVMVWKKG 303  
 320 TERHHPGF---EATFQPLPMS--SCGRLAKAGCTFSPYFGHYPPNIDCTWNIIEVPN 374  
 304 MHSYDPEFLSVKVAFPQDCQVNLTLBGRLL-DTQGFRTPPYSPYSSTHCSMHLTVPS 362  
 375 NQHVKVPKFFYLLPEGVPACTCPDVEIENGKYGCR-----SQFVTSNSMKITVRF 429  
 363 LDYGLALWFDAVALRQCKNRLCTQGGMTQNRRLCGRTTQPAERLIPWASDGVITNF 422  
 430 HSDSYDTDTGLAEYLSYDSDPCPGQPTCTGTCIRKELCDGADCTHSDIELNSCD 489  
 423 TSQSLSGPQVQVYVYSLYNSDPCGEBLGSV----- 454  
 490 AGHOFCKKCKCKELFWVCDSVNDGDSDEQSGSCPAQTQRC-SNGCKLSKQCCNKD 548  
 455 -----NGLCVP---ACDGIKDCPCMLDNRNCVCRAM-FOCEDSTCISLPRVCROP 502  
 549 DCGSGDSBASPKNVNVCTGHTYRCLNGLCLSKGNPCDCKGDCSDSDKDCDCCGLRS 608  
 503 DCLNGSDEBQOE--GVPCGFTTQCEBRSCVKKPNPCDQSDSDRDSQHCDCUQCG 560  
 609 FTROARVVGCTDADEGEWPMQVSLHALQCGHICGASLISPMWLVSAAHCYIDRGRYS 668  
 561 LS--SRIVGTVSSEGEWPMQASLQIRNR-HICGALILADRWVITAAHCFQED---SWAS 614  
 669 PTQWTAFLGLHDGOSRSPVQERLKAIIISHPPFNFTFYDIALLEKPAVSSNVR 728  
 615 PKLWTVFLGKMRQNSR--WPGEVSEFVSRLLPLHPYHEBDSHBYDVALQLDHPVVSATVR 673  
 729 PICLPDASHVPAGKAVVWTGHTQVGTGALILQKEIRIVNQTGENTLLPOQTPRM 788  
 674 PVCLPANSHPFEPQCHWITGMAGOREGSPVSNLQKVDVQVLPQDLCSKARYQVSPRM 733  
 789 MCVFSLSGVSDSCGDSGGLPSVEADRIIGAGVSWGDCAGRNKPEVYTRLPFLRDM 848  
 734 LCAVYRKGKXKACGDSGGLPVCREPSRWFLAGLVSGLCGRNPFQVTVTRVAVNW 793  
 849 IKE 851  
 794 IQQ 796

NCBI\_Taxid=10090;  
 [1]  
 SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.  
 STRAIN=C57BL/6J;  
 MEDLINE=22755759; PubMed=12744720; DOI=10.1042/BJ20030390;  
 RA Hooper J.D., Campagnolo L., Goodarzi G., Truong T.N., Stuhlmann H.,  
 RA Outgley J.P.;  
 RT "Mouse matrilase-2: identification, characterization and comparative  
 RT mRNA expression analysis with mouse hepsin in adult and embryonic  
 RT tissues.";  
 RL Blochem. J. 373:689-702(2003).  
 [2]  
 SEQUENCE FROM N.A.  
 RP STRAIN=C57BL/6J; TISSUE=Liver;  
 RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nigaki I., Osato N., Saito R., Suzuki H., Yamana I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schirral L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chochia C., Corbett L.E., Cousins S.,  
 RA Dalla B., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Petter G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sadelain A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takeda Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Wetanabe Y., Wells C.,  
 RA Wilming L.G., Wyszynski-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Zaninetti P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yaunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Bitney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 [3]  
 SEQUENCE FROM N.A.  
 RP TISSUE=Liver;  
 RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
 RA Hopkins R.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Altschul S.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stetler-Stevenson M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Tothiyaki S., Giannini P., Prange C.,  
 RA Bata S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McSwan P.U., McKernan K.J., Malek J.A., Guneratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Skalska D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Matra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [4]  
 REVIEW.  
 RX MEDLINE=22668120; PubMed=12784999;  
 RA Netzel-Arnett S., Hooper J.D., Szabo R., Madison E.L., Outgley J.P.,  
 RA Bugge T.H., Antalis T.M.;



AC O81U80; O81U82; O81XV8;  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Transmembrane protease, serine 6 (EC 3.4.21.-) (Matricinase-2).  
 GN Name="TMPRSS6";  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Hooper J.D., Quigley J.P.;  
 RT "TMPRSS6, a new type II transmembrane serine protease";  
 RL Submitted (SFP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208; DOI=10.1038/990031;  
 RA Dunham I., Hunt A.R., Collins J.E., Brunkiewicz R., Beare D.M.,  
 RA Clamp M., Smink L.J., Atencough R., Almeida J.P., Babbage A.K.,  
 RA Begganley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
 RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
 RA Gilbert J.G., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,  
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
 RA Laird G.K., Langford C.F., Levermore M.A., Lloyd C., Lloyd D.M.,  
 RA Martin I.D., Maheirgh-Mohammadi M., Matthews L.H., Mccann O.T.,  
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,  
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L.,  
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
 RA Soderlund C., Spragon L., Steward C.A., Suleston J.E., Swann R.M.,  
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilting L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Minoshima S., Kawaasaki K., Sasaki T., Aakawa S., Kudoh J.,  
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenyon S., Lai H., Lao H.I.,  
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malt J.E., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shauli S., Sloan D., Song L.,  
 RA Wang Q., Wang Y., Wang Z., White Z., Willingham D., Wu H., Yao Z.,  
 RA Zhan M., Zhang G., Chisose S., Murray J., Miller N., Minx P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rohlfing T.,  
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
 RA Korf I., Bebell J.A., Hillier L.W., Mardis E., Waterston R.,  
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,  
 RA Budarf M.L., Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
 RA Edelman L., Kim U.J., Shizuya H., Simon M.I., Dumack J.P.,  
 RA Peyrard M., Keda D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,  
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
 RA Khan A.S., Lane L., Tilihan Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22";  
 RL Nature 402:489-495(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner K.H., Shanmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carinini P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villard D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bonfield G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywnski M.I., Skalska U., Smalhus D.E.,  
 RA Scherach A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP SEQUENCE OF 9-811 FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.  
 RC TISSUE=Testis;  
 RX MEDLINE=22241917; PubMed=12149247; DOI=10.1074/jbc.M203007200;  
 RA Velasco G., Cai S., Quesada V., Sanchez L.M., Lopez-Otin C.;  
 RT "Matricinase-2, a membrane-bound mosaic serine proteinase predominantly  
 RT expressed in human liver and showing degrading activity against  
 RT extracellular matrix proteins";  
 RL J. Biol. Chem. 277:37637-37646(2002).  
 RN [5]  
 RP REVIEW.  
 RX MEDLINE=22668120; PubMed=12784999;  
 RA Netzel-Annett S., Hooper J.D., Szabo R., Madison E.L., Quigley J.P.,  
 RA Bugge T.H., Anstie T.M.;  
 RT "Membrane anchored serine proteases: a rapidly expanding group of cell  
 RT surface proteolytic enzymes with potential roles in cancer";  
 RL Cancer Metastasis Rev. 22:237-258(2003).  
 CC -1- FUNCTION: May play a specialized role in matrix remodeling  
 CC processes in liver.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=O81U80-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=O81U80-2; Sequence=VSP 008379, VSP 008380;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: Liver specific.  
 CC -1- SIMILARITY: Belongs to peptidase family 11.  
 CC -1- SIMILARITY: Contains 2 CUB domains.  
 CC -1- SIMILARITY: Contains 3 LDL-receptor class A domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AJ319876; CAC85953.1; ALT\_INIT.  
 CC EMBL; AY055383; AAL16413.1; -  
 CC EMBL; AY055384; AAL16414.1; -  
 CC EMBL; AL022314; -; NOT ANNOTATED\_CDS.  
 CC EMBL; BC039082; AA039082.1; -  
 CC HSSP; P00760; 1BZX.  
 CC Genew; HGNC:16517; TMPRSS6.  
 CC InterPro; IPR000859; CUB.  
 CC InterPro; IPR002172; LDL\_receptor\_A.  
 CC InterPro; IPR001254; Peptidase\_S1.  
 CC InterPro; IPR009003; Pept\_Ser\_Cys.  
 CC Pfam; PF00431; CUB; 1.  
 CC Pfam; PF00057; Ldl\_recept\_a; 2.  
 CC Pfam; PF00089; Trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC PRINTS; PR00261; LDLRECEPTOR.  
 CC SMART; SM00042; CUB; 1.  
 CC SMART; SM00192; Lda; 3.  
 CC SMART; SM00020; TRYP\_SPC; 1.  
 CC PROSITE; PS01180; CUB; 1.

DR		PROSITE; PS01209; LDLRA_1; 1.
DR	PROSITE;	PS02068; LDLRA_2; 3.
DR	PROSITE;	PS0240; TRYPSIN_DOM; 1.
DR	PROSITE;	PS00134; TRYPSIN_HIS; 1.
DR	PROSITE;	PS00135; TRYPSIN_SER; 1.
KW	Alternative splicing;	Glycoprotein; Hydrolase; Repeat;
	Serine protease; Signal-anchor;	Transmembrane.
FT	DOMAIN	1 55
FT	TRANSMEM	56 76
FT	DOMAIN	77 811
FT	DOMAIN	213 336
FT	DOMAIN	335 452
FT	DOMAIN	457 489
FT	DOMAIN	490 526
FT	DOMAIN	530 567
FT	DOMAIN	577 811
FT	ACT SITE	617 617
FT	ACT SITE	668 668
FT	ACT SITE	762 762
FT	CARBOHYD	136 136
FT	CARBOHYD	184 184
FT	CARBOHYD	216 216
FT	CARBOHYD	338 338
FT	CARBOHYD	433 433
FT	CARBOHYD	453 453
FT	CARBOHYD	518 518
FT	VARSPLIC	409 461
FT		
FT		
FT		
FT	VARSPLIC	462 811
FT		
FT	CONFLICT	116 116
SO	SEQUENCE	811 AA; 89999 MW; 75EFD193F655DDE9D CRC64;

Query	March	23.8%	Score 1114:	DB 1:	Length 811:
Best	Local Similarity	31.0%	Pred.	Mo. 2.3e-68:	
Match	276:	Conservative 136:	Mid.	No. Matches 329:	Indels 148: Gaps 28:
QY	11	GQPKDFAGLKYNSHREKYNGLSEVEELPVNNVKKVKGHGGRVNVLAAYLIGLLVYL	70		
DB	20	GGGGGGGGG-----EEAPEGMFKACE-----DSKRRARGRLRPVFPVLL--ALTIVA	66		
QY	71	GIGFLVMH-LQYR-DYRQYKXNGMYR.TTNEFVDAYENSNSTEVSILASKYKDLXLY	128		
DB	67	SAGVLMFLFKGKAEVWVSQYISGSLKRLNRRFSODLTRRESARSRSETAQAQKMLKELT	126		
QY	129	SGVPLFGLPYHESAAVTAFSQSVLIVYVSEFSIPOLVLEAEARVAEERVVALPEPRAEKL	188		
DB	127	TSTR-LGTYYNSSSVYSFGEGPLTFCFWFIQIDPH-----RLMLSPREV--V	171		
QY	189	KSFVTVTSVVAAPPTDSKYQRTQDNCSGSLARGVLEMRF-----TPRGFDSPY-	238		
DB	172	QALTEBELLSTVNSSAAPVPRAE-----EYVDEGVLIVLEASYKDIALNSTGTCTRYSYV	227		
QY	239	-----PAH--ARCOMALRGDADSVLSLTFPSFDIASCDEGSDLVLYNYLTSME	286		
DB	228	GQGVQLRLKGPDPHLASSCLMHLQGPDKDMLKRLR-EMTLAEGRDR-----LAWYDVAGPLE	282		
QY	287	PHALVOL--CGTYPPTSYMLT-----FHSSQVLLILITLTERRHPEFATP	331		
DB	283	KRLITSYVGGCRQDEVEVLELSSGALIMAVWKKGLHSYDPPFLASV-----QPVV	331		
QY	332	FQLPRMSSCGRLRKAGQSTFNSPYRGHPYPNIDTVMNILEVNNOHVAKVRFKFFYLEBPG	391		
DB	332	FQACQVNLTLNRLRDSQGVLSPTFYPYSYSPQTHCSMWLTPVSLDYGALMADAAALARKQ	391		
QY	392	VPAGTCFMDIYVEINGEKYCG-----ERSQFVTVTSNSKTIYVAFHSDQSYTDGFLAE	443		
DB	392	KYDLPCTQGWMTIQNRLCGRLILQPAERILPVAVTAG-----ITINFTSQISLTGGVAVH	448		

[illegible]

RESULT 13		
ID	Q6UXD8	PRELIMINARY; PRT; 802 AA.
AC	Q6UXD8:	
DT	05-JUL-2004 (TREMBLrel. 27, Created)	
DT	05-JUL-2004 (TREMBLrel. 27, Last sequence update)	
DT	05-JUL-2004 (TREMBLrel. 27, Last annotation update)	
DE	PVAE354.	
CN	ORFNames=UNQ354;	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
OX	NCBI_TaxId=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=22887296; PubMed=12975309;	
RA	Clark H.F., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,	
RA	Chen J., Chow B., Chui C., Crowley C., Curriel B., Deuel B., Dowd P.,	
RA	Bacon D., Foster J., Grimaldi C., Gu Q., Haas P.E., Heldens S.,	
RA	Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,	
RA	Lewis L., Liao D., Mark M., Robble E., Sanchez C., Schoenfeld J.,	
RA	Seethagiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,	
RA	Vandlen R., Watanabe C., Wiand D., Woods K., Xie M.H., Yansura D.,	
RA	Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,	
RA	Goddard P.;	
RT	"The secreted protein discovery initiative (SPDI), a large-scale	
RT	effort to identify novel human secreted and transmembrane proteins: a	
RT	bioinformatics assessment";	
RL	Genome Res. 13:2265-2270 (2003).	
CC	-1- SIMILARITY: Belongs to peptidase family S1.	
DR	EMBL; AY358398; AAC88764.1; -.	
DR	InterPro; IPR000859; CUB.	
DR	InterPro; IPR002172; LDL_receptor_A.	
DR	InterPro; IPR001254; Peptidase_S1.	
DR	InterPro; IPR001314; Peptidase_S1A.	
DR	Pfam; PF0049003; Pept_ser_Cys.	
DR	Pfam; PF004311; CUB; 1.	
DR	Pfam; PF00057; LDL_recept_a; 2.	
DR	Pfam; PF00089; Trypsin; 1_.	
DR	PRINTS; PR00722; CHYMOTRYPSIN.	
DR	PRINTS; PR00261; LDLRECEPTOR.	
DR	SMART; SM00192; LDLA; 3.	
DR	SMART; SMO0020; Tryp_spec; 1.	
DR	PROSITE; PS01180; CUB; 1.	



DR PROSITE; PS01209; IDLRA\_1; 1.  
 DR PROSITE; PS50068; TRYPSIN\_DOM; 3.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 802 AA; 88845 MW; 8726C91B1E02E163 CRC64;

Query Match 23.7%; Score 1112; DB 2; Length 802;  
 Best Local Similarity 31.0%; Pred. No. 3.1e-68;  
 Matches 276; Conservative 135; Mismatches 330; Indels 148; Gaps 28;

11 GGPDPFAGLKYNRHKEVGLGVEFLPVNNVKKYKGGKGRWVLAVALIGLLVLL 70  
 11 GGGDGGDG-----EEAEPGEMFAC-----DSKRAKGYLRVLVPLVLL--ALLVLA 57  
 QY 71 GIGFLVNH-LQYR-DVRQVKYKGYMRITNENFVDAVENSSTEFVSLASKVDALKLY 128  
 DB 58 SAGVLWVFLGYKAEVWVSQVYSGSLVLRNHFQDGLTRRESSAFRSETAKAQMKLELI 117  
 QY 129 SGVFLGPRYHKEAVNTAFSGSVIAYVSEPSIPQHLVEAEKRVMAERVMLEPPRARS 188  
 DB 118 TSTR-LGTYYNSSSVYSGEGPLTCFEMFIIQIPEH-----RRLMLSPV--V 162  
 QY 169 KSPVYTSVVAFPDTSKYVQTQDNSSCSFGLHARGVELMR-----TTGPPDSPY- 238  
 DB 163 QALVEBELSTVNSSAAVPRAB-----YVDPEGLVILEASVKDIALNSTLGCYRYSYV 218  
 QY 239 -----PAH--ARCOMALRGDADSVLSLTFRSPDLASCDERSDLVTVVNTLSFME 286  
 DB 219 GGGVLRKPGDHLASSCLMHLQGPDKMLKRL--EWTLACGR--LAMYVAGGLE 273  
 QY 287 PHALVQL--CCTYPPSYNLT-----FHSSGNVLLITLITNTERRHGFEATF 331  
 DB 274 KRLITSYVSGSRQRPVVEVLASGAIMAVWVKGLHSYDPEFLSV-----QPVV 322  
 QY 332 FQLRMSSCGRLKKAOGTFNSPYRGHYPPNIDCTNIEVPNNQHVKVRKFFYLEPG 391  
 DB 323 FQACEVNLTLNRLDSQGVLSLTFPSYSPQTHCSMHLTVPSLDYGLAMFDAYALRRQ 382  
 QY 392 VPAGTCPRDYVEINGEKYCG-----ERSQPVVTSNSNKITVRFHSDQSYDTGFLAE 443  
 DB 383 KYDLPTCGQWNTIGNRLCGRLIQPYAERLIPVATAG---ITINFSQISLITGPGVRVH 439  
 QY 444 YLSYDSDPCPGQGTCTGTGRCIRKELRCGMACTDHSDELNSCDAHGQFTCKNKKCP 503  
 DB 440 YGLYNOSDPCEGFLCSV-----NGLCVP 463  
 QY 504 LFWVCDSVNDCGNSDEGSCSPAQTFRG--SNKGLSKSQOQNGKDDCGDSDASCPKV 562  
 DB 464 --ACDGVKDCPNGLDERNCVCRATFQCKEDSTCISLPKVCDCQPDCLNSDEQCE- 518  
 QY 563 NVVYCTKTATYRCLNGLCLSKGNPECDGKEDSCDEKDCDGLRSFTQRARVVGSTAD 622  
 DB 519 -GVCGTTFPCGDBRSYKKNPCQDGRPRDCDSDDEHCCGLOG--PSSRIYGANSS 575  
 QY 623 EGEPMQVSLHALCGSHICGASLISPMVLVAANHCTYIDRGFRYSDDPTQWTAFLGHDQS 682  
 DB 576 EGEPMQVSLHALCGSHICGASLISPMVLVAANHCTYIDRGFRYSDDPTQWTAFLGHDQS 682  
 QY 683 QRSAPGVGERLKLKLIISHPFNDFTFDVIALLEKFAEYSSWVRPCLDASHVPRAG 742  
 DB 632 SR-WPEGVSVFVSRLHLPHYHEBDSHDYVALLDLDHVVRSAAVRPCLPARSHFEPG 690  
 QY 743 KAIWVTCGHTQYGGTALILQKEIRIVINQTCENLLPOQITPRMCKVGLSGGVSDSC 802  
 DB 691 LHCWITGKALRBEGPISNALQKYDVQLIPQDLCEAVRYQVTPRMLCAGRKKGKACQ 750  
 QY 803 GDSGGLSSVADGRIFGAGVYVSWDGCAGKNGKGVYTRLPFRDWIK 851  
 DB 751 GDSGGLPVCALSGRWFLAGLVSWGLCGGRNYRGVYTRITGVISWIOQ 799

## RESULT 14

ID: AA088764 PRELIMINARY; PRT; 802 AA.  
 AC AA088764;  
 DT 02-MAR-2004 (TREMBlrel. 27, Created)  
 DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
 DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)  
 DE PVAE354.  
 GN UNQ354.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primata; Catarrhini; Homnidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=12975309;  
 RA Clark H.F., Gureley A.L., Abaya E., Baker K., Baldwin D., Brush J.,  
 RA Chen J., Chow B., Chui C., Crowley C., Currell B., Deuel B., Dowd P.,  
 RA Baton D., Foster J., Grimaldi C., Gu O., Hase P.B., Heldens S.,  
 RA Huang A., Kim H.S., Kijowski L., Jin Y., Johnson S., Lee J.,  
 RA Lewis L., Liao D., Mark M., Robbie E., Sanchez C., Schoenfeld J.,  
 RA Seehagiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,  
 RA Vandlen R., Watanabe C., Wiedand D., Woods K., Xie M.H., Yansura D.,  
 RA Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,  
 RA Godowski P.;  
 RT "The Secreted Protein Discovery Initiative (SPDI), a Large-Scale  
 RT Effort to Identify Novel Human Secreted and Transmembrane Proteins: A  
 RT Bioinformatics Assessment."  
 RL Genome Res. 13:2265-2270(2003).  
 DR EMBL; AY358398; AA088764.1; -  
 SQ SEQUENCE 802 AA; 88845 MW; 8726C91B1E02E163 CRC64;

Query Match 23.7%; Score 1112; DB 2; Length 802;  
 Best Local Similarity 31.0%; Pred. No. 3.1e-68;  
 Matches 276; Conservative 135; Mismatches 330; Indels 148; Gaps 28;

11 GGPDPFAGLKYNRHKEVGLGVEFLPVNNVKKYKGGKGRWVLAVALIGLLVLL 70  
 11 GGGDGGDG-----EEAEPGEMFAC-----DSKRAKGYLRVLVPLVLL--ALLVLA 57  
 QY 71 GIGFLVNH-LQYR-DVRQVKYKGYMRITNENFVDAVENSSTEFVSLASKVDALKLY 128  
 DB 58 SAGVLWVFLGYKAEVWVSQVYSGSLVLRNHFQDGLTRRESSAFRSETAKAQMKLELI 117  
 QY 129 SGVFLGPRYHKEAVNTAFSGSVIAYVSEPSIPQHLVEAEKRVMAERVMLEPPRARS 188  
 DB 118 TSTR-LGTYYNSSSVYSGEGPLTCFEMFIIQIPEH-----RRLMLSPV--V 162  
 QY 169 KSPVYTSVVAFPDTSKYVQTQDNSSCSFGLHARGVELMR-----TTGPPDSPY- 238  
 DB 163 QALVEBELSTVNSSAAVPRAB-----YVDPEGLVILEASVKDIALNSTLGCYRYSYV 218  
 QY 239 -----PAH--ARCOMALRGDADSVLSLTFRSPDLASCDERSDLVTVVNTLSFME 286  
 DB 219 GGGVLRKPGDHLASSCLMHLQGPDKMLKRL--EWTLACGR--LAMYVAGGLE 273  
 QY 287 PHALVQL--CCTYPPSYNLT-----FHSSGNVLLITLITNTERRHGFEATF 331  
 DB 274 KRLITSYVSGSRQRPVVEVLASGAIMAVWVKGLHSYDPEFLSV-----QPVV 322  
 QY 332 FQLRMSSCGRLKKAOGTFNSPYRGHYPPNIDCTNIEVPNNQHVKVRKFFYLEPG 391  
 DB 323 FQACEVNLTLNRLDSQGVLSLTFPSYSPQTHCSMHLTVPSLDYGLAMFDAYALRRQ 382  
 QY 392 VPAGTCPRDYVEINGEKYCG-----ERSQPVVTSNSNKITVRFHSDQSYDTGFLAE 443  
 DB 383 KYDLPTCGQWNTIGNRLCGRLIQPYAERLIPVATAG---ITINFSQISLITGPGVRVH 439  
 QY 444 YLSYDSDPCPGQGTCTGTGRCIRKELRCGMACTDHSDELNSCDAHGQFTCKNKKCP 503  
 DB 440 YGLYNOSDPCEGFLCSV-----NGLCVP 463  
 QY 504 LFWVCDSVNDCGNSDEGSCSPAQTFRG--SNKGLSKSQOQNGKDDCGDSDASCPKV 562



Db 464 ---ACDVKKCPNGLDERNCVCRA-TFOCKEDSTCISLPKVCQGPDLNLSGSEBQOE- 518  
Qy 563 NVVTCGTHYKTLNGLCLSKNPECDGKEDCDSDKDCGRLSRSTRQARVVGTDAD 622  
Db 519 -GVPCGFTTFQCEDRSCVKKENPQCDGRPDGRSDDEHDCGLQG--PSSRTVGGAVSS 575  
Qy 623 EGEMPWQVSLHALGQGHICGASLISPMWLVSAAHCYIDRGRFYSDEPTQTAFLGHDQS 682  
Db 576 EGEMPWQVSLALQVRGR-HICGGLIADRWVITTAHCFQED--SWASRVLMVTLFGKWMQN 631  
Qy 683 QRAPGVQERLRKRIISHPFNDFTFYDIALLEKPARVSSMVRICLPDASHVPAG 742  
Db 632 SR-WPGVSPKVSRLILHPHEBDSHDYVALQLDHPVRSAAVRPVCCLPARSHFEPG 690  
Qy 743 KAIWVTGMGHTQYGGTGAALIQKEIRVINGTTCENLLPQOITPRMVCVGLSGVNSCQ 802  
Db 691 LHCHITGMGLRREGGPIISNALQKVYDVLIPDOLCSAERYQVTPRMLCAGYRKCKDQACQ 750  
Qy 803 GDSGGLPSVYADGRIFGAGVSWGDCAGRNKPYVTRLPFRDWIKX 851  
Db 751 GDGSGPLVCALSGRWFLAGLVSGLGCGRPNGYVTRITGVISWIOQ 799

## RESULT 15

Q61CC2 PRELIMINARY; PRT; 824 AA.  
ID 061CC2  
AC 061CC2;  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, last annotation update)  
DE D01170K4.2 protein.  
GN Name=D01170K4.2;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Collins J.E., Wright C.L., Edwards C.A., Davie M.P., Grinham J.A.,  
RA Cole C.G., Goward M.E., Aguado B., Mallya M., Mokrab Y., Huckle E.J.,  
RA Beare D.M., Dunham I.,  
RL Submitted (May-2004) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to peptidase family S1.  
DR EMBL; CR456446; CAG30332.1; -.  
DR InterPro; IPR000859; CUB.  
DR InterPro; IPR002172; LDL\_receptor\_A.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR Pfam; PF00431; CUB; 1.  
DR Pfam; PF00057; LDL\_recept\_a; 2.  
DR Pfam; PF00089; Trypsin; 1.  
DR PRINTS; PRO0722; CHYMOTRYPSIN.  
DR PRINTS; PRO0261; LDLRECEPTOR.  
DR SMART; SM00192; LDla; 3.  
DR SMART; SM00020; TRYP\_SPC; 1.  
DR PROSITE; PS01180; CUB; 1.  
DR PROSITE; PS01209; LDla\_2; 1.  
DR PROSITE; PS50068; LDla\_2; 3.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KM Hydrolyase; Protease; Serine protease.  
SQ SEQUENCE 824 AA; 91333 MW; A74F186406041F7B CRC64;

Query Match 23.3%; Score 1092; DB 2; Length 824;

Best Local Similarity 30.3%; Pred. No. 7.8e-67;

Matches 276; Conservative 135; Mismatches 330; Indels 170; Gaps 29;  
Qy 11 GGPDPGAGKYNNRHKVNGLEGVFELPVNNVKKYKYEKGPRGWVYLAVALIGLLVYL 70  
Db 11 GGQGGGSDG-----EEAPPEGMFVACE-----DSKRKANGYRLVLEFVILL--ALLVLA 57

Qy 71 GIGFLVWH-LQYR-DYRVQKYNKYNRITNENFVDAENSNSTEFVSLASKVADLKLY 128  
Db 58 SAGVLLMFLYGYAEVWVSQVSGSLRVLNRRHSQDLTRRESAFAFSEPAKQMKELI 117  
Qy 129 SGVFPFGPHKKEAVALYAFSEGSVIAIYVSEFSIPQILVEAEVVAEERVMLPPARSL 188  
Db 118 TSTR-LGYTNSSSVYSFEGGPLTCFFWFLIQPEH-----RRLMISPEV--V 162  
Qy 189 KSFVYVSVAFPDSTKYVORTODNSCSFGHARGVILMF-----TTPGFPDSPY- 238  
Db 163 QALIVEELSTVNSSAAVPRAB-----YEVDEGLVILASVYDIAANSTLGCTYSYV 218  
Qy 239 -----PAH--ARQWALRGDADSVLSLTFRSFDLASCDERSDLVTYNTLSPEE 286  
Db 219 GQGVRLRKGPDLHASSCLMHLQGPDKMLKRL-EMTLAECDR-----LAMYDVAGPLE 273  
Qy 287 PHALVOL--CGTYPBYNLT-----FHSSQVLLITLTNTERRRPGEATP 331  
Db 274 KRLITVYGCGRQEPVEVLAAGAIMAVVKKGLHSYDBFVLSV-----QPVV 332  
Qy 332 FQLPKMSGGRKRAQGTENSPYRPHYVNPINDCTWNIENPNNOHVKRFKFFYLLERQ 391  
Db 323 FQACEVNLIDNRLDSQVLSLTFPFSYSPQTHCSMHLTVESLDYGLALMPAYALRQ 382  
Qy 392 VPAGTCQVVEINGERYCG-----ERSQFVYTSNKNKTVRFHSDQSYDTGFLAE 443  
Db 383 KYDLPTQGWNTIONRLGLRLIQPYARIPIVATAG--ITINFTSQSLTGPGRVYH 439  
Qy 444 YLSYDSDPQPGPFTRTGRCIRKELRCDGMADCTHSDBLNCSCAHQFTCKNFQXP 503  
Db 440 YGLYNSDPQPGFELSV-----NGLCYV 463  
Qy 504 LFWVCSVNDGNSDEGCSCPAOTFRCSNKGCLSKSQQCNKGKDCDGDSDDEASCPTY 562  
Db 464 ---ACDGVKDCPGLDERNCVCRA-TFOCKEDSTCISLKVCDGQDCLNLSGSEBQOE- 518  
Qy 563 NVVTCGTHYKTLNGLCLSKNPECDGKEDCDSDKDCGRLSRSTRQARVVGTDAD 622  
Db 519 -GVPCGFTTFQCEDRSCVKKENPQCDGRPDGRSDDEHDCGLQG--PSSRTVGGAVSS 575  
Qy 623 EGEMPWQVSLHALGQGHICGASLISPMWLVSAAHCYIDRGRFYSDEPTQTAFLGHDQS 682  
Db 576 EGEMPWQVSLALQVRGR-HICGGLIADRWVITTAHCFQED--SWASRVLMVTLFGKWMQN 631  
Qy 683 QRAPGVQERLRKRIISHPFNDFTFYDIALLEKPARVSSMVRICLPDASHVPAG 742  
Db 632 SR-WPGVSPKVSRLILHPHEBDSHDYVALQLDHPVRSAAVRPVCCLPARSHFEPG 690  
Qy 743 KAIWVTGMGHTQYGGTGA-----LILQKEIRVINGTTCENLL 780  
Db 691 LHCWITGMGLRREGALRADVALFYGMRNQSEFTCCPISNALQKVYDVLIPDOLCSERV 750  
Qy 781 PQOITPRMVCVGLSGVNSCQDSDGGLSVADGRIFGAGVSWGDCAGRNKPYVTR 840  
Db 751 RYGVTPRMLCAGYRKCKDQACQDGGPLVCALSGRWFLAGLVSGLGCGRPNGYV 810  
Qy 841 RLPLFRDWIKX 851  
Db 811 RITGVISWIOQ 821

## RESULT 16

CAG30332 PRELIMINARY; PRT; 824 AA.  
ID CAG30332  
AC CAG30332;  
DT 01-JUN-2004 (TREMBlrel. 27, Created)  
DT 01-JUN-2004 (TREMBlrel. 27, last sequence update)  
DT 01-JUN-2004 (TREMBlrel. 27, last annotation update)  
DE D01170K4.2 protein.  
GN Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,  
 RA Cole C.G., Goward M.E., Aguado B., Mallya M., Mokrab Y., Huckle E.J.,  
 RA Beare D.M., Dunham I.,  
 RA Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CR456446; CAG30332.1;  
 SQ SEQUENCE 824 AA; 91333 MW; A74F186406041F7B CRC64;  
 Query Match 23.3%; Score 1092; DB 2; Length 824;  
 Best Local Similarity 30.3%; Pred. No. 7,8e-67;  
 Matches 276; Conservative 135; Mismatches 330; Indels 170; Gaps 29;  
 QY 11 GGPRDFAGLKYKSRHEKNGLEGEFLPVNNVKKVEKHPGKRWVLAVALIGLLVLL 70  
 DB 11 GGQGDGDDG-----EEAEPEEGMFACR-----DSKRKARGYLRLVPLFVLL--ALVLA 57  
 QY 71 GIGLVMH-LOYR-DYAVOKKXNGVMRTNENFVDAYENSTREFFVSLASKVDAKALVLL 128  
 DB 58 SAGVLLWYFLGKYKAEVMSQVYSGSLRVLNHFQODLTRRESSAFRSETAAQKMLLELI 117  
 QY 129 SGVPLGPKYHKSAYTAFSEGSVLAAYWSEFIPQHLVEAEKRYMAEERVMLEPPRARSLL 188  
 DB 118 TSTR-LGTYNASSSVYSGEGPLTCFPIQIDPH-----RLMLSPV--V 162  
 QY 189 KSPVVTSVVAPEPTDSKYVQRTODNSCSFGLHARGVELMR-----TTPGPPDSY- 238  
 DB 163 QALVEEILSTVNSSAAVPYRAE---YEVDPEGLVLEASVKDIAMALNSTLGCYRYSYV 218  
 QY 239 -----PAH--ACQWALRGDADSVSLTFRSPFLASCDERSGLVTVVNTLSPE 286  
 DB 219 GQGVLRLLKGDHLASSCLWHLQPKDMLKRL-EWTLACRUR---LAMYVAGLE 273  
 QY 287 PHALVOL--CGTYPSPYVLT-----FHSSQNVLLTLITTERHHPGEATF 331  
 DB 274 KRLITSYVCGSRQRPVEVLASGAIMVWKKGLHSHYDPPVLV-----QPVV 322  
 QY 332 FQLRMSSCGRLRKAQGTSPYPGHYPPNIDCTNIEVNNQHVKKRFFYLLPBG 391  
 DB 323 FOACEVNLTLDRDLSQGVSLTPYFPYSYSPQTHCSWMLTVPSLDYGLALFPAVALRLQ 382  
 QY 392 VPAGTCRKDYVEINGEKCG-----ERSQFVTSNSMKTITRPHSDQYTOTGLAE 443  
 DB 383 KYDLPCTQGWNTNNRRLCGRLILOPAERLPVAVATAG--ITINFQSISLTSPGARVH 439  
 QY 444 YLSYDSDPCEGQFTCTGRCIRKELRCGDWADCTDHSDELINSCDAGHQTCKKFKCP 503  
 DB 440 YGLYNQSDPCRGELCSV-----NGLCVP 463  
 QY 504 LFWVCDVNDGSDNSDEQSCSPAQTFRC-SNGKCLSKSQCGNGKDDGSDSDEASCPKV 562  
 DB 464 ---ACDGVKDCPNGLDERNCVCRA-TFOCKEDSTCISLPKCDQDQDLNCSDEQCOE- 518  
 QY 563 NVVTCTKATYKCLNGLCLSKNPECDGDESDSDSDCCGGLRSFTRQARVVGCDAD 622  
 DB 519 -GVPCGFTFQCEBRSCVKKPNQCDGPRDCRDSDBEHCQCGQG--PSSRIYGAVSS 575  
 QY 623 EGEWPMQVSLHALCGAHLISPMVLVSAACYIDRFRYSDDPTQWAFGLDQDS 682  
 DB 576 EGEWPMQASLOVRGR-NICGALIALDRKIVITAHACFGQD---SNASTYLWVFLGKQWQN 631  
 QY 683 QRSAPGQVQERLKAIIHPFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVPAG 742  
 DB 632 SR-WPGEVSPKYSRLLHPYHEBDSHDYVALLQLDHPVVSAAVRPCLPARSHFEPG 690  
 QY 743 KAIWVTGNGHTQYGTGA-----LILQKGRVNVNQTTCENLL 780  
 DB 691 LHCWITGALREGLRLDAVALPYGWNQSEITCCCPISNALQKVDQVLLPQDLCESEVY 750  
 QY 781 PQQTPRMKCVGLSGVDSCGDSGGLPVSVEADGRIFGAGVSWMGDCAGRRKPGVYT 840

DB 751 RYQVTPMLCAGYRKGDACQDSDGGLVCKALSGRWFLAGVSMGLGCGRPYFGVYT 810  
 QY 841 RLPLFRDWKE 851  
 DB 811 RITGVISWIDQ 821  
 RESULT 17  
 ID Q7RTY8 PRELIMINARY; PRT; 572 AA.  
 AC Q7RTY8  
 DT 01-MAR-2004 (TREMBLrel. 26, Created)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)  
 DT 01-OCT-2004 (TREMBLrel. 28, Last annotation update)  
 DE Type II transmembrane serine protease 7 precursor (hypothetical  
 protein FLJ16086).  
 GN Name=TMPRSS7;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22722134; Pubmed=12838346;  
 RA Puente X.S., Sanchez L.M., Overall C.M., Lopez-Otin C.,  
 RT "Human and mouse proteases: a comparative genomic approach."  
 RL Nat. Rev. Genet. 4:544-558 (2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Nimomoya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,  
 RA Katsura N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kamai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isegai T.,  
 RA Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 RL -I- SIMILARITY: Belongs to peptidase family S1.  
 CC EMBL; BN000125; CAD67577.1; -.  
 DR EMBL; AKJ31211; BAD18401.1; -.  
 DR MEROPS; S01.072; -.  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0008223; F:peptidase activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR001859; CUB.  
 DR InterPro; IPR002172; LDL\_receptor\_A.  
 DR InterPro; IPR001254; peptidase\_S1.  
 DR InterPro; IPR001314; peptidase\_S1A.  
 DR Pfam; PF00431; CUB; 1.  
 DR Pfam; PF00057; Ldl\_recept\_a; 2.  
 DR Pfam; PF00089; Trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYSIN.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR PROSITE; PS01180; CUB; 2.  
 DR PROSITE; PS01209; LDLRA\_1; 1.  
 DR PROSITE; PS01068; LDLRA\_2; 2.  
 DR PROSITE; PS02440; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM HydroLase; Protease; Serine protease; Signal; Transmembrane.  
 FT SIGNAL 1 23  
 FT CHAIN 24 572 type II transmembrane serine protease 7.  
 SQ SEQUENCE 572 AA; 64041 MW; 0458CBEA96EA285 CRC64;  
 Query Match 21.4%; Score 1004.5; DB 2; Length 572;  
 Best Local Similarity 36.6%; Pred. No. 5.6e-61;  
 Matches 228; Conservative 95; Mismatches 223; Indels 77; Gaps 23;  
 QY 244 COMALRGDADSVSLTFRSPFLASCDERSGLVTVVNTLSPEPHALVOLCGTYPSPYVLT 303

Db 2 CHEKDAVAVGLYLRSLRSIKSIQI-EADNCVDTSLTIYDSLPIRSLILYRIC-EPTRTLM 58  
Qy 304 TEPSSQVLLITLTITNTERRHPEFATEFOLPRMSQSGRLRK-----AOGFPNSPYRPH 359  
Db 59 SFTSTNNLVTFSPHRIIRLSGRARFERYLPBCKCENTVAVKQITGPEKISSPYPSY 118  
Qy 360 YPPNIDCTWNIIEVNNQHVKRFKPF-YLLEPGVPACTCPDYVEINGEKYCGE--RSQF 416  
Db 119 YPCKCKCTWKEQT-SLSTLGIALKFYNYSTIKSMKG-CEHGMEWEINBHMYCGSYMDHQ 176  
Qy 417 VVTSNKKITVRFPSIDSYDTDTGLAELYSDDPCP-GQFTCRGRCKRKEICRGMA 475  
Db 177 IFRVPSLVHLQLQCSSRLSDKPLAAYGSYNISQPCVPSPFRCSGICVPAQRCGVN 236  
Qy 476 DCTHPSDELNCSCDAGHQFTCKNFC---KPLFWVCDSVNDGDNDEGSCSPADOTFC 532  
Db 237 DCFPESBEL-----FCVSPQP-----ACTTSFR- 260  
Qy 533 SNGKLSKSQOCNGKDCGDSDBASCPKVVVCTRHRYCLNGCLSKGNEPCDGED 592  
Db 261 QHGPLI-----CDGFRCENGDRQNC--TQSI PCNNRTFCGNDICFRKQNAKCDGTVD 313  
Qy 593 CSQSDSKDCDCGKRSRTQRAVVGTDADGEMPMQVSLHALQCGHICGASLISPMVLV 652  
Db 314 CPDSDDEGCTCS-RSSSALHRIIGGDTLEGGMVQVSLHFGVSAY-CGASVISREML 371  
Qy 653 SAACHYIDDRGFRYSDPTQMTAFGLHDGORS--APGVERRRLKRIISHPFNDPFDDY 710  
Db 372 SAACHF---HGNRISDPTPTAHGMYQGNKAVSP-----VRKLVHETVNSQTFDY 422  
Qy 711 DIALLELE--KPAEYSSWVRPILCPDASHVFPACKAIWVGWHT-OYGGTGAIILOKGE 767  
Db 423 DIALQLGSIAMPETLKQILQIPICIPPGQVRVSGEKCMTVGWRHBDKNGSLVLQGA 482  
Qy 768 IRVINGTTCENLRFQOTTFRMCMVGFUSGVDSCQSGSGLS-SVADGRIFGAYVSW 826  
Db 483 VELIDQTLGVSTY-GIITSRLCGAGISGRDACKDGSGLSGRRKDGKMLITGLVSW 541  
Qy 827 GDGCAGRNKPGVYTRPLPLFRDWI 849  
Db 542 GHGGRNPGVYTRVSNFVPMI 564

RESULT 18  
Q8BIK6 PRELIMINARY; PRT; 572 AA.  
AC Q8BIK6;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:B230219123 product:weakly similar to BLOOD COAGULATION FACTOR XI.  
DE Name=B230219123Rik;  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=99279253; PubMed=10349636;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Meth. Enzymol. 303:19-44(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=21085660; PubMed=11217851;  
RA RIKEN FANTOM Consortium;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [3]

RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RA The FANTOM Consortium.  
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=2049374; PubMed=11042159;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes";  
RL Genome Res. 10:1617-1630(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=20530913; PubMed=11076861;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Kono H., Akiyama J., Nishi K., Kitsuai T., Taishiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M., Oneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RA Aichi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P., Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K., Hiraoka I., Hirozane T., Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T., Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M., Nishi K., Nomura K., Numazaki R., Ohno M., Ohseio N., Okazaki Y., Saito R., Saitoh H., Sakai C., Sakai K., Sakazune N., Sano H., Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M., Tagawa A., Takahashi F., Takaku-Akahira S., Takada Y., Tanaka T., Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
DR EMBL, AK045663; BAC32448.1; -.  
DR HSP, P00760; IEZX.  
DR MEROPS, S01.072; -.  
DR MGD, MGI:2686594; B230219123Rik.  
DR GO, GO:0004263; F:chymotrypsin activity; IEA.  
DR GO, GO:0004295; F:trypsin activity; IEA.  
DR GO, GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro, IPR000859; CUB.  
DR InterPro, IPR002172; LDL\_receptor\_A.  
DR InterPro, IPR001254; Peptidase\_S1.  
DR InterPro, IPR001314; Peptidase\_S1A.  
DR InterPro, IPR009003; Pept\_Ser\_Cys.  
DR Pfam, PF00431; CUB.1.  
DR Pfam, PF00057; Ldl\_recept\_a; 3.  
DR Pfam, PF00089; Trypsin.1.  
DR PRINTS, PR00722; CHYMOTRYPSIN.  
DR PRINTS, PR00261; LDLRECEPTOR.  
DR SMART, SM00042; CUB.1.  
DR SMART, SM00192; LDBa; 3.  
DR SMART, SM00020; TYP\_Spc; 1.  
DR PROSITE, PS01180; CUB.2.  
DR PROSITE, PS01209; LDLRA\_1; 1.  
DR PROSITE, PS50068; LDLRA\_2; 2.  
DR PROSITE, PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE, PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE, PS00135; TRYPSIN\_SER; 1.  
DR Hydrolase; Protease; Serine protease.

SQ SEQUENCE 572 AA; 63757 MW; 21C0EC523B5F4301 CRC64;  
 Query Match 21.4%; Score 1004.5; DB 2; Length 572;  
 Best Local Similarity 36.5%; Pred. No. 5.6e-61;  
 Matches 226; Conservative 94; Mismatches 229; Indels 71; Gaps 21;  
 QY 244 COMALRGDADSVLSLTFRSPDLASCDERGSDLTVYNTLSPMEMHALVOLCGTPSPSYNL 303  
 DB 2 CHFLVAVIGVYLRLSISIQI-EADNCITDLSLTVDSLPIRSAILYRIC-EPTRTLM 58  
 QY 304 TFFSSQVLLITLITNTERHPGEATFPLPRMSSCGG-----RLKAOCTFNSPYYPG 358  
 DB 59 SFVSTNNMLVTLKSPYRLAGIRAYFEVPE-QKCESTLIVKINSPEKISSPYPS 117  
 QY 359 HYPNIDCTWNIENVNNOHVVRKFFYLEPGVPAGTCPPDYVEINKEKCGE--RSQF 416  
 DB 118 YPPKCKCTWTFQT-SLSTGLAKFYVYSITKKSACEGHEWINEHMYCGSYMDHET 176  
 QY 417 VVTSNKKITVRFRHSDSYTDGFLAEYLSYDSDPCP-GQFTCRGRCIKELRCDGMA 475  
 DB 177 IFRVPSPLVHIQLCCSRRLSDKPLAEYGSYNSQCCAGSFRCSGGLCVPAQRCDGVN 236  
 QY 476 DCTDHSDELNCSGADGHOFTCKNFKCPLFWVCDSDVNDGDNDSDEGSCSPAQTFRC 535  
 DB 237 DCFPESDELFCVT-----VXP-----ACNSSSFR-QHG 263  
 QY 536 KCLSKSQOQNGKDDCGSDSEASCPRKNVVTCTKHTYRCLNGCLSKNPECDGKEDCD 595  
 DB 264 PLV-----CGFRDCBEGDEQONCTP-SIPCTSRTEFCGNDICFRKQNAQCDGIVDCPD 316  
 QY 596 GSDSKDDCGLRSTROARVVGTDADGEMPMQVSIHALGQGHICGASLISPMNLVSA 655  
 DB 317 GSDEEGCGCS-RSSSFLHRIYGSDDSGTWPQVSLHFGVSAV-CGASVISREMLLSAA 374  
 QY 656 HCYIDDRGFRYSPTQTWTAFLGLHDSQSR--APGVERRLKRIISHPFNDFDYDIA 713  
 DB 375 HCF---HGNRLSDPTPTTAHLGMVVOGNAKFISSP-----VRRLVHEVYNSQTFDDIA 425  
 QY 714 LLELE--KPAEYSSMVRPCLPDASHVFPAGKAIWVGWHT-QYGTGALLILOGEIRV 770  
 DB 426 LQLQSIAMPETLKLQIOLIPICIPAGOKVRSGEKCMTGWRHEDSKSGSVLQDAEVEL 485  
 QY 771 INQTCENLIPQOITPRPMCVGLSGVDSGCGSGLS-SVEADGRIFGAGVNSG 829  
 DB 486 IDQVVCSTY-GITTSRLCAGVSGSKSDACKDSGGLSCRRKSDGKWLITGLVSWHG 544  
 QY 830 CAGRNKPGVYTRLELPRDWI 849  
 DB 545 CGRNPFGVYTRVSSVFWPI 564  
 RESULT 19  
 BAD18401 PRELIMINARY; PRT; 572 AA.  
 AC BAD18401:  
 DT 12-MAY-2004 (TrEMBLrel. 27, Created)  
 DT 12-MAY-2004 (TrEMBLrel. 27, Last sequence update)  
 DE CDNA FLJ16088 f18, clone NT2RP7008435, weakly similar to EPITHTIN (EC 3.4.21.-).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX Mammalia\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nimomiya K., Magatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Futuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,  
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato K., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,

RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.,  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (MAR2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK31211; BAD18401.1;  
 SQ SEQUENCE 572 AA; 64041 MW; 0458CBEA96BA285 CRC64;  
 Query Match 21.4%; Score 1004.5; DB 2; Length 572;  
 Best Local Similarity 36.6%; Pred. No. 5.6e-61;  
 Matches 228; Conservative 95; Mismatches 223; Indels 77; Gaps 23;  
 QY 244 COMALRGDADSVLSLTFRSPDLASCDERGSDLTVYNTLSPMEMHALVOLCGTPSPSYNL 303  
 DB 2 CHFLVAVIGVYLRLSISIQI-EADNCITDLSLTVDSLPIRSAILYRIC-EPTRTLM 58  
 QY 304 TFFSSQVLLITLITNTERHPGEATFPLPRMSSCGGLR-----AQCTFNSPYYPG 359  
 DB 59 SFVSTNNMLVTLKSPYRLAGIRAYFEVPEYPOKCENTLVXDITGFEKISSPYPSY 118  
 QY 360 YPPNIDCTWNIENVNNOHVVRKFF-YLLEPGVPAGTCPPDYVEINKEKCGE--RSQF 416  
 DB 119 YPPKCKCTWTFQT-SLSTGLAKFYVYSITKKSMTG-CEGHEWINEHMYCGSYMDHQT 176  
 QY 417 VVTSNKKITVRFRHSDSYTDGFLAEYLSYDSDPCP-GQFTCRGRCIKELRCDGMA 475  
 DB 177 IFRVPSPLVHIQLCCSRRLSDKPLAEYGSYNSQCPVGSFRCSGGLCVPAQRCDGVN 236  
 QY 476 DCTDHSDELNCSGADGHOFTCKNFKC--KPLFWVCDSDVNDGDNDSDEGSCSPAQTFRC 532  
 DB 237 DCFPESDELFCVT-----FCVSPQ-----ACNTSFR- 260  
 QY 533 SNKCLSKSQOQNGKDDCGSDSEASCPRKNVVTCTKHTYRCLNGCLSKNPECDGKED 592  
 DB 261 QHGELI-----CGFRDCBEGDEQONCTP-TQSPCNKRTFCGNDICFRKQNAQCDGTV 313  
 QY 593 GSDSKDDCGLRSTROARVVGTDADGEMPMQVSIHALGQGHICGASLISPMNLV 652  
 DB 314 CPDSDEGCTCS-RSSSALHRIIGGDTLEGMPQVSLHFGVSAV-CGASVISREML 371  
 QY 653 SAACHYIDDRGFRYSPTQTWTAFLGLHDSQSR--APGVERRLKRIISHPFNDFDY 710  
 DB 372 SAACHF---HGNRLSDPTPTTAHLGMVVOGNAKFSVSP-----VRRLVHEVYNSQTFDY 422  
 QY 711 DIALLELE--KPAEYSSMVRPCLPDASHVFPAGKAIWVGWHT-QYGTGALLILOGE 767  
 DB 423 DIALQSIAMPETLKLQIOLIPICIPAGOKVRSGEKCMTGWRHEDSKSGSVLQDAE 482  
 QY 768 IRVINOCTCENLIPQOITPRPMCVGLSGVDSGCGSGLS-SVEADGRIFGAGVNS 826  
 DB 483 VELIDQTLVSTY-GITTSRLCAGVSGSKRDKDSGGLSCRRKSDGKWLITGLVSW 541  
 QY 827 GDCGAGRNKPGVYTRLELPRDWI 849  
 DB 542 GHGCGRNPFGVYTRVSNFVFWPI 564  
 RESULT 20  
 ENTX\_PIG STANDARD; PRT; 1034 AA.  
 AC P98074;  
 DT 01-FEB-1996 (Ref. 33, Created)  
 DT 01-FEB-1996 (Ref. 33, Last sequence update)  
 DT 05-JUL-2004 (Ref. 44, Last annotation update)  
 DE Enteropelliclase precursor (EC 3.4.21.9) (Enterokinase).  
 GN Name=PRSS7; Synonyms=ENTX;  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sue.  
 OX Mammalia\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RA TISSUE=Duodenal mucosa;  
 RX MEDLINE=94327548; PubMed=8051081;  
 RA Matsushima M., Ichinose M., Yahagi N., Kakei N., Tsukada S., Miki K.,



```

Db 196 SIECLPSRPAADLAKCIAVLDLFCDDGELNCPDSDSDSKICATACD-----GKFLILTES 249
Qy 230 TPFGPDSEYP-----AHARQWALRGDADSVSLTFRSGDLASCER-----GSDLV--- 276
Db 250 SGSEFDAQYPLTSEASVVCQWIIIRVNOGLSIELNFSYFNTYSMVULNIEGVGSSKILRA 309
Qy 277 -----TVVNTLS----- 283
Db 310 SLMLNPETIRIFSNQVTVTLFLISDENDYIGFNATYAFNSTELNDEKINCPEDEFC 369
Qy 284 -----PM 285
Db 370 FWIGDLNDNEMERIQGTFPPFGPNPDHTEFGNASGYISTPGPGRQROEVGLSLSLPL 429
Qy 286 EPHALVOLC-----GYPPSYN----- 302
Db 430 EP-TLEPVCLSFWYMYGENVYKLSINISNDQNEKIIFOKEGVYGENWYNGVTLNETV 488
Qy 303 ---LTFHSSQVLLITLITNTERHHPF-----EATFF-----QLPRMSSCGG--RLR 345
Db 489 EFKAFFAFKQFSLDALDDISLTYYGICNVSLYEPFLVPTSPPELP--TDCGAPPELW 546
Qy 346 KAQGFNSPYYPGHYPPNIDCTNANIEVNNQHKVRFKFFYLEPBGVPACTCPKDYVEI- 404
Db 547 EPNTTFSMNPNNYPNAPFCVMNLNOKKGIQLHPEFLENI-----DVEIR 598
Qy 405 NGEK-----YCGERSQFVNTSNNKITYRFHSDSYDTGFLAELYS---YDSSDP 453
Db 599 DGEEDSLILLAVYGPGEVDVFTTNMTVLFTINDALTGKGFANFTTGYHLGIDEP 658
Qy 454 -PGQFTGRTGRCIKELACDGMADCTHSDDELNCSADGHPCTCKNKRCKLFWVCSYN 512
Db 659 KEDNFOCGNCECVLLVNLCDGFSHCKGSDAHACV-----RF-----LN 697
Qy 513 DCGNSDEQCGSCPAQTFRC---NGKCLSKSQCCNGKDC---GDGSDASCP----- 560
Db 698 GTANNSG-----LVQFRIQSIWHTACAEWTTQTSIDVQQLGLGTGNSMFPFSSCG 750
Qy 561 ---KVVVVTCTKTYRCLNGLCLSKGNPECDGKED-----CSDGSDKDCDGLSFT 610
Db 751 GPVFKLNTAP-----NGSLILITASBQC--FEDSLILLCQNHKS-----CGKQVA 793
Qy 611 RQA--RVVGGTADABEGMPWQVSLHALGQGHICGASLISPMWLYSAACHYIDDDGFRYSD 668
Db 794 QEVSPKIVGNDSEGAWPVVALYVNGQ--LLCGASLVSRLMVLVSAACHYV---RWLE 848
Qy 669 PTQNTAFGLHDORSAPGVQERRLKRIISHPFNDFPDYDIALLEKPAEYSSMVR 728
Db 849 PSKKKAILGLHMTSNLTSPOIVTRLIDETIVNPHYNNRRKDSDIAMHMLEFKVNTYDQ 908
Qy 729 PICLPDASHVPPAKAIWVTGMHTQYGGTALILQKGEIRVINQTCENLLPO-QITPR 787
Db 909 PICLPBENQVPPRISCSIAGMGKVIYQGSPPADILQEADVPLLSNEKQOQMPENITEN 968
Qy 788 MNCVGFSLSGVDSQGSQSGPLSSVEADGRIFGAGVYVSWGCGCAKRNKPGVYTRLPLFRD 847
Db 969 MMCAGYEBGIGISQGSQSGPLMCLL--NNRWLAGVYSFGYQALPNRPVYAAVPEKTE 1027
Qy 848 WIK 850
Db 1028 WIK 1030

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=22784742; Pubmed=12886014;
RA Cal S., Lopez-Otin C.;
RT "Polymerase-I, a human polypotease with the ability to generate
RT independent serine protease domains from a single translation
RT product.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:9185-9190(2003).
DR EMBL; AJ488947; CAD35759.1; -.
DR MEROPS; S01.357; -.
DR MEROPS; S01.358; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001272; LDL_receptor_A.
DR InterPro; IPR001254; peptidase_S1.
DR Pfam; PF00057; Ldl_recept_a; 1.
DR Pfam; PF00089; Trypsin; 2.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00020; TRYP_SPC; 2.
DR PROSITE; PS50068; LDLa_2; 1.
DR PROSITE; PS50240; TRYPIN_DOM; 2.
DR PROSITE; PS00134; TRYPIN_HIS; UNKNOWN_2.
DR PROSITE; PS00135; TRYPIN_SER; 2.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 855 AA; 91352 MW; 8AF2759D9740CF3F CRC64;

Query Match 15.2%; Score 711.5; DB 2; Length 855;
Best Local Similarity 45.7%; Pred. No. 1.8e-40;
Matches 134; Conservative 52; Mismatches 96; Indels 9; Gaps 5;

Qy 567 CTKTYRCLNGLCLSKGNPECDGKEDCSDGSDKDCDGLRSPTRQ-RVVGTDADGE 625
Db 154 CPNPSFCGNSQCVTKNPKPCDDQEDCSDGSDAHCEGCLQPMARMAGRIVGVGWSA 213
Qy 626 WPMQVSLHALGQGHICGASLISPMWLYSAACHYIDDRGFYSPDTQNTAFGLHDORS 685
Db 214 FPMQASLIRE-NKEHFGAALINARMLVSAACHYNE-----FODPTKVAVYAGTYLSGE 267
Qy 686 APGVQERRLKRIISHPFNDFPDYDIALLEKPAEYSSMVRPICLPDASHVPPAKAI 745
Db 268 ASTYR-AQVQIVGHPVYNADFDVAVLELTSPLPFGHIDPVCLPAATHIFPESKCC 326
Qy 746 WVTGMHTQYGG--TGALLIQKGEIRVINQTCENLLPQITPRMVCVGFSGVDSQGD 804
Db 327 LISGMVGLKDLPLVKREVLQKATVELLDQALCASLYGHSLTDRMVCAGYLDGKVDSCGD 366
Qy 805 SGGPLSSVEADGRIFGAGVYVSWGCGCAKRNKPGVYTRLPLFRDIKENT 853
Db 387 SGGPLVCEBPSGRFFLAGIVSWGIGCAEARPGVYAVTRLRDMILEAT 435

RESULT 22
Q72411 PRELIMINARY; PRT; 1059 AA.
ID Q72411
AC Q72411;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Polymerase-Ia protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=22784742; Pubmed=12886014;
RA Cal S., Lopez-Otin C.;

```

RT "Polypease-I, a human polypeptidase with the ability to generate  
RT independent serine protease domains from a single translation  
RT product."  
RL Proc. Natl. Acad. Sci. U.S.A. 100:9185-9190(2003).  
DR EMBL: AJ488946; CAD35758.1; -.  
DR MEROPS; S01.357; -.  
DR MEROPS; S01.358; -.  
DR MEROPS; S01.969; -.  
DR GO; GO:0004295; F:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR002172; LDL\_receptor\_A.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR Pfam; Pf00057; Ldl\_recept\_a; 1.  
DR Pfam; Pf00089; Trypsin; 3.  
DR SMART; SM00192; Ldlra; 1.  
DR SMART; SM00020; Tryp\_spc; 3.  
DR PROSITE; PS50068; LDLRA\_2; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 3.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN; 3.  
DR PROSITE; PS00135; TRYPSIN\_SER; 2.  
KM Hydrolase; Protease; Serine protease.  
SQ SEQUENCE 1059 AA; 114020 MW; 17D27A2D99F2A264 CRC64;

Query Match 15.2%; Score 711.5; DB 2; Length 1059;  
Best Local Similarity 45.7%; Pred. No. 2.3e-40;  
Matches 132; Conservative 52; Mismatches 96; Indels 9; Gaps 5;

QY 567 CTKHYRCLNGLCISKGNPCDGDGDEKDCGCGLRSPFROA-RVVGCTDADEGE 625  
DB 154 CPGNSFCGSCQCTKYNPEDEDDSDSDENHCEGCLGAPARMAGRVGMEAPGE 213  
QY 626 WPMQVSLHALGQGHICGASLISPMVLVAHCTYIDRGFRYSPTQWTAFLGHDGQRS 685  
DB 214 PPMQASLRE-NKEHFCAALINAMVLVAHCFNE-----FQDPTKWAVYGATYLSGE 267  
QY 686 APQGERLKRILHPFNPTPYDIALLEKPAVSSWVRICLPDASHVPACKAI 745  
DB 268 ASTVR-AQVAVQIVGHPYMDAPDAVALVETLSPFGRIHQVCLPAATHIPPSKCC 326  
QY 746 WVTSGMTOYGG-IGALILKGKEIRVINTTCENILPQQLTPRMVCGVSGVDSQGD 804  
DB 327 LISGMGLIKEDFLVKEPVQKATYELDQALCASLHSHSLTDRAVCGAGYLDGKSDSCGD 386  
QY 805 SGGPLSSVEADGRIFGAGVSWDGCAGRNKPGVYTRLPFRMIXENT 853  
DB 387 SGGPLVCEPSPGRFFLAGIVSWGICAEARRPGYAVTRLDWILBAT 435

RESULT 23  
ENTR BOVIN STANDARD; PRT; 1035 AA.  
ID ENTR BOVIN STANDARD; PRT; 1035 AA.  
AC P98072;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).  
GN Name=PRSS7; Synonyms=ENTK;  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovinae; Bos.  
OC NCBI\_TaxID=9913;  
RX 11  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Duodenum;  
RX MEDLINE=94329561; PubMed=8052624.  
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;  
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic  
RT protease composed of a distinctive assortment of domains."  
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).  
RN [2]  
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=94043122; PubMed=8226855;

RA Lavallie E.R., Rehmetulla A., Racie L.A., DiLaasio E.A., Ferenz C.,  
RA Grant K.L., Light A., McCoy J.M.;  
RT "Cloning and functional expression of a cDNA encoding the catalytic  
RT subunit of bovine enterokinase."  
RL J. Biol. Chem. 268:23311-23317(1993).  
RN [3]  
RP SEQUENCE OF 801-827.  
RC TISSUE=Intestine;  
RX MEDLINE=92189715; PubMed=1799406;  
RA Light A., Janska H.;  
RT "The amino-terminal sequence of the catalytic subunit of bovine  
RT enterokinase."  
RL J. Protein Chem. 10:475-480(1991).  
CC -1- FUNCTION: Responsible for initiating activation of pancreatic  
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase  
CC A). It catalyzes the conversion of trypsinogen to trypsin which in  
CC turn activates other proenzymes including chymotrypsinogen,  
CC procarboxypeptidases, and proelastases.  
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in  
CC trypsinogen.  
CC -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a  
CC multidomain (heavy) chain linked by a disulfide bond.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=Long;  
CC IsoId=P98072-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=P98072-2; Sequence=VSP\_005386;  
CC -1- TISSUE SPECIFICITY: Intestinal brush border.  
CC -1- PFM: The chains are derived from a single precursor that is  
CC cleaved by a trypsin-like protease.  
CC -1- SIMILARITY: Belongs to peptidase family S1.  
CC -1- SIMILARITY: Contains 2 CUB domains.  
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.  
CC -1- SIMILARITY: Contains 1 SEA domain.  
CC -1- SIMILARITY: Contains 1 SEA domain.  
CC -1- SIMILARITY: Contains 1 SRCR domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: U09859; AAB40026.1; -.  
DR EMBL: L19663; AAI16035.1; -.  
DR PIR: A43090; A43090.  
DR PDB; 1EKX; X-ray; A=788-800, B=801-1035.  
DR MEROPS; S01.156; -.  
DR InterPro; IPR000859; CUB.  
DR InterPro; IPR002172; LDL\_receptor\_A.  
DR InterPro; IPR000998; MAM.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR InterPro; IPR011163; Pept\_S1A\_enterop.  
DR InterPro; IPR009003; Pept\_Set\_Cys.  
DR InterPro; IPR000082; SEA.  
DR InterPro; IPR001190; Srcr\_receptor.  
DR Pfam; Pf00431; CUB; 2.  
DR Pfam; Pf00057; Ldl\_recept\_a; 2.  
DR Pfam; Pf00629; MAM; 1.  
DR Pfam; Pf01390; SEA; 1.  
DR Pfam; Pf00530; SRCR; 1.  
DR Pfam; Pf00089; Trypsin; 1.  
DR PIRSF; PIRSF001138; Enteropeptidase; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00261; LDLRECEPTOR.  
DR PROSITE; PS01180; CUB; 2.  
DR PROSITE; PS01209; LDLRA\_1; 2.  
DR PROSITE; PS50068; LDLRA\_2; 2.



DR	PROSITE; PS00740; MAM 1; 1.	
DR	PROSITE; PS50060; MAM_2; 1.	
DR	PROSITE; PS50024; SEA; 1.	
DR	PROSITE; PS00420; SRCR 1; FALSE_NEG.	
DR	PROSITE; PS50287; SRCR 2; 1.	
DR	PROSITE; PS50240; TRYPsin DOM; 1.	
DR	PROSITE; PS00134; TRYPsin HIS; 1.	
DR	PROSITE; PS00135; TRYPsin_SER; 1.	
KM	3d-structure; Alternative splicing; Direct protein sequencing;	
KM	glycoprotein; Hydrolyase; Lipoprotein; Myristate; Repeat;	
KM	Serine protease; Signal-anchor; Transmembrane; Zymogen.	
FT	CHAIN 1 800	
FT	CHAIN 801 1035	
FT	DOMAIN 1 18	
FT	TRANSMEM 19 47	
FT	DOMAIN 48 1035	
FT	DOMAIN 54 169	
FT	DOMAIN 197 238	
FT	DOMAIN 240 330	
FT	DOMAIN 358 520	
FT	DOMAIN 540 650	
FT	DOMAIN 657 695	
FT	DOMAIN 694 787	
FT	DOMAIN 801 1035	
FT	ACT_SITE 841 841	
FT	ACT_SITE 892 892	
FT	ACT_SITE 987 987	
FT	LIPID 2 2	
FT	DISULFID 199 212	
FT	DISULFID 206 225	
FT	DISULFID 219 226	
FT	DISULFID 659 671	
FT	DISULFID 666 684	
FT	DISULFID 678 693	
FT	DISULFID 788 912	
FT	DISULFID 826 842	
FT	DISULFID 926 932	
FT	DISULFID 957 972	
FT	DISULFID 983 1011	
FT	CARBOHYD 116 116	
FT	CARBOHYD 147 117	
FT	CARBOHYD 170 170	
FT	CARBOHYD 194 194	
FT	CARBOHYD 233 233	
FT	CARBOHYD 263 263	
FT	CARBOHYD 264 264	
FT	CARBOHYD 404 404	
FT	CARBOHYD 456 456	
FT	CARBOHYD 486 486	
FT	CARBOHYD 519 519	
FT	CARBOHYD 550 550	
FT	CARBOHYD 646 646	
FT	CARBOHYD 698 698	
FT	CARBOHYD 722 722	
FT	CARBOHYD 741 741	
FT	CARBOHYD 762 762	
FT	CARBOHYD 864 864	
FT	CARBOHYD 903 903	
FT	CARBOHYD 965 965	
FT	VARSPLIC 166 192	
FT	CONFLICT 808 808	
FT	STRAND 802 802	
FT	STRAND 805 806	
FT	TURN 809 810	
FT	TURN 813 814	
FT	STRAND 815 820	
FT	TURN 821 822	
FT	STRAND 823 830	
FT	STRAND 835 838	
FT	HELIX 840 843	
FT	TURN 844 845	
R -> Y (in Ref. 3).		

[illegible]

```

Db      543 PHDLWENTTFTSINFPNSYPNQAFCIMWNAOKGXIOHFOFDEJENIA-----D 594
Qy      401 YVEI-NEKTCGERSOV-----TYSNKNKTIVFPHSQSTYDGLAEYLS-- 446
Db      595 VVEIRDSB---GDDSLFLAYITGPGPVNDVFTSTNMTVLEFIDNMLAKQGFANFTTGY 651
Qy      447 -YDSDEPC-PGQFTCRTRGCIKRLKRCGMADCTDHSDELNC-----SCDAGH--OPTC 496
Db      652 GLGIPBCKEDNFOCKDGECLPLVNLCDGPHCKDGSDEAHCVRLFNGTTDSSGLVGFRI 711
Qy      497 KAKFCRKLFWVCDSVNDCCGNSBQSGSCAQTFRCSNGKCLSSQCGNCKDQDGSDE 556
Db      712 QS-----IMHV-----ACAEN-----MTTQISDDVC-----QLLG-----LGTNSS 743
Qy      557 ASC-----PRVNVVTCFKATYRCNLGLCLSKGNPECDKEDCSGSDKDCD---CGLR 607
Db      744 VPTFTSGGPGPVNLTNP-----NCSLITPEQC-----LEDSLILIQCNKSCGCK 791
Qy      608 SFTROA--RVVGTDADEGEWPMQVSLHALGQCHIICASLISPMLVSAACHYIDRGR 665
Db      792 LVTOGEVSPKIVGSDSREGAMPWVALYPDDQ--QVCGASLVSRDMLVSAACHCYG---R 846
Qy      666 YSDPTQMTALGLHDQGRAPGVQERLKRILSHPFNDFTPYDIALLEKPAEYSS 725
Db      847 NMESKMKAVLGLHMAENLTSPOLETLIDQIVINPHYNRKRKNNDIAMMILEKRVYTD 906
Qy      726 MVRPICLPDASHVPAKAIWVTGMGHTQYGTGALILQKGEIRVINQTCENLLPQ-QI 784
Db      907 YIQCIPCEPNQVFPFRICISIAQMGALITYQSTADVLQADVLSNEKCOQOMPEYNI 966
Qy      785 TPRMVCVPLSGVSDSCQDSGSGPLSSVEADGRIFGAGVSWDGCAGRRKPGVYTEL 844
Db      967 TENNVCAGYBAGVSDSCQDSGSGPLMCOE--NNRWILAGVTSFGVQCALPWRPGYAVR 1025
Qy      845 FRDMIK 850
Db      1026 FTEWIQ 1031

```

```

RA      Van W., Wu F., Morser J., Wu Q.;
RT      "Corin, a transmembrane cardiac serine protease, acts as a pro-atrial
RT      natriuretic peptide-converting enzyme.";
RL      Proc. Natl. Acad. Sci. U.S.A. 97:8525-8529 (2000).
CC      -1- FUNCTION: Converts Pro-ANP to ANP. Cleaves Pro-ANP specifically
CC      between Arg-123 and Ser-124.
CC      -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC      -1- TISSUE SPECIFICITY: Highly expressed in heart. Expressed in heart
CC      myocytes.
CC      -1- SIMILARITY: Belongs to peptidase family S1.
CC      -1- SIMILARITY: Contains 7 fingered (FZ) domains.
CC      -1- SIMILARITY: Contains 1 SCR domain.
CC      -1- SIMILARITY: Contains 1 SCR domain.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL: AP133845; AAD31850.1; -.
DR      EMBL: AP113248; AAP21966.1; -.
DR      HSSP: P01130; IAUU.
DR      MEROPS: S01.019; -.
DR      Genew: HGNC:19012; CORIN.
DR      MIM: 605236; -.
DR      GO: GO:0005887; C: integral to plasma membrane; TAS.
DR      GO: GO:0004252; F: serine-type endopeptidase activity; TAS.
DR      GO: GO:0006629; P: lipid metabolism; TAS.
DR      GO: GO:0006653; P: morphogenesis; TAS.
DR      GO: GO:0006508; P: proteolysis and peptidolysis; TAS.
DR      GO: GO:0008217; P: regulation of blood pressure; TAS.
DR      InterPro: IPR000024; FZ domain.
DR      InterPro: IPR002172; LDL_receptor_A.
DR      InterPro: IPR001254; Peptidase_S1.
DR      InterPro: IPR001314; Peptidase_S1A.
DR      InterPro: IPR009003; Pept_Ser_Cys.
DR      InterPro: IPR001190; Srrc_receptor.
DR      Pfam: PF01392; Fz_2.
DR      Pfam: PF00057; Ldl_recept_a; 6.
DR      Pfam: PF00089; Trypsin; 1.
DR      PRINTS: PR00722; CHYMOTRYPSIN.
DR      PRINTS: PR00261; LDLRECEPTOR.
DR      PROSITE: PSS0038; FZ_2.
DR      PROSITE: PSS01209; LDLRA_1; 6.
DR      PROSITE: PSS0068; LDLRA_2; 7.
DR      PROSITE: PSS0420; SRRc_1; FALSE_NEG.
DR      PROSITE: PSS0287; SRRc_2; FALSE_NEG.
DR      PROSITE: PSS0240; TRYPSIN_DOM; 1.
DR      PROSITE: PSS0134; TRYPSIN_HIS; FALSE_NEG.
DR      PROSITE: PSS0135; TRYPSIN_SER; 1.
KW      Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor;
KW      Transmembrane.
KX      DOMAIN 1 45
FT      TRANSMEM 46 66
FT      DOMAIN 67 1042
FT      DOMAIN 134 259
FT      DOMAIN 268 304
FT      DOMAIN 305 340
FT      DOMAIN 341 377
FT      DOMAIN 378 415
FT      DOMAIN 450 573
FT      DOMAIN 579 614
FT      DOMAIN 615 653
FT      DOMAIN 654 690
FT      DOMAIN 690 786
FT      DOMAIN 802 1042
FT      ACT_SITE 843 843
FT      ACT_SITE 892 892
FT      ACT_SITE 985 985

```

RESULT 24  
CORI\_HUMAN STANDARD; PRT; 1042 AA.  
AC Q9Y5Q5; Q9UHY2;  
DT 16-OCT-2001 (Ref. 40, Created)  
DT 16-OCT-2001 (Ref. 40, Last sequence update)  
DT 05-JUN-2004 (Ref. 44, Last annotation update)  
DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-  
DE converting enzyme) (Corin) (Heart specific serine proteinase ATC2).  
GN Name=CORIN; Synonyms=CRN;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart;  
RX MEDLINE=99262646; PubMed=10329693;  
RA Van W., Sheng N., Seto M., Morser J., Wu Q.;  
RT "Corin, a mosaic transmembrane serine protease encoded by a novel cDNA  
RT from human heart.";  
RL J. Biol. Chem. 274:14926-14935 (1999).  
RN [2]  
RP SEQUENCE OF 734-1040 FROM N.A.  
RC TISSUE=Heart;  
RX MEDLINE=20534769; PubMed=11082206;  
RA Hooper J.D., Scarnan A.L., Clarke B.E., Normyle J.F., Antalis T.M.;  
RT "Localization of the mosaic transmembrane serine protease corin to  
RT heart myocytes.";  
RL Eur. J. Biochem. 267:6931-6937 (2000).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=20359740; PubMed=10880574; DOI=10.1073/pnas.150149097;

```

FT DISULFID 790 912 By similarity.
FT DISULFID 828 844 By similarity.
FT DISULFID 955 970 By similarity.
FT DISULFID 981 1010 By similarity.
FT CARBOHYD 80 80 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 104 104 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 135 135 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 141 141 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 231 231 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 245 245 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 251 251 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 305 305 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 320 320 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 376 376 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 413 413 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 446 446 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 451 451 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 469 469 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 567 567 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 651 651 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 697 697 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 761 761 N-linked (GlcNAc . . .) (Potential).
FT MUTAGEN 1022 1022 N-linked (GlcNAc . . .) (Potential).
FT CONFLICT 854 854 S->A: Loss of activity.
FT CONFLICT 876 876 W->R (in Ref. 2).
SQ SEQUENCE 1042 AA, 116564 MW, 7705398B8607AD2 CRC64;

Query Match 14.8%; Score 695; DB 1; Length 1042;
Best Local Similarity 28.2%; Pred. No. 3,1e-39;
Matches 210; Conservative 97; Mismatches 265; Indels 172; Gaps 32;

Qy 202 DSKTVQRTQDNCSGFLHARGVELMRTTGPFPDPSYPFAHARQOMALRGDAD----- 253
Db 365 DHDCVDSDEYNCS--CHSQGLVECR-----NSCQIPSTRQCC-----GDEDCXGDSDE 412
Qy 254 --SVLSTLFRSFD-----LASCDERGSPLTVVNTL-----SPMEPHALVOLCGTYPPS 300
Db 413 NCSYIQTSQCGDRCLYNPLDSCG--GSPLCPNNSLNCSQCEPTTL-ELCMNLP-- 467
Qy 301 YNLTF-----HSSQNVLLITLINTERRHGFATFQPLRMSSCGRLKKAQGTNSP 354
Db 468 YNYSYNYNFGHRQKASIS-----WESSLFPALVQNYCYKYLMPFECTI--- 513
Qy 355 YYPGHYPNIDCTNINIEVPN---NQYKVFKEFYLLPEVPAGTCPKQVYEINGEKYC 410
Db 514 ----LVPKCDVNTGERIIPCRALCEHSKERCEVSL-----GIVGLQWPEDE----- 554
Qy 411 GERSQFVNTSNKRTVRFHSDQSYTDGFLAEVLSYDSSDPC--PGQFTGTCGRCIKEL 469
Db 555 TDCSQFPEBNSDQTCI-----MPDEVY-----ECSFSHPKCRSGQCVLASR 597
Qy 470 RCDMACTDHSDBLANSCDAGHOFTC--KNFKCKPLFVCDVNDGDNDEQCS--CPA 527
Db 598 RCDGQACDDSDDENCGCKRDLMECPNSNCKLKHVYICGFPDCCPYMDKKNKSCFQD 657
Qy 528 QTFRCNSGKCLSKSQCCNGKDDCGDSDASCPKVN----- 564
Db 658 DELECAHACVSRDLMDCEADGSDSDSDMDCVLSINVSSEFLMVRATEHHVACDG 717
Qy 565 -----VTC-----TKHT--YKCLNGCLSK-----GNPBDG 589
Db 718 WOELISQACQKMGLEBPSYTKLIOEOKERPMULTLSHNSMISLGTTHLLVNGQSCS 777
Qy 590 KEDSDSDKDCDCGRLSFTFR--QARVVGTDADEGEVPMQVSLHALGQGHICASLISP 648
Db 778 RSKSLSLCTKQ--DQGRPARAMNKRILGRTSRGRRPWCSCLOSBSHGICCVILAK 835
Qy 649 NMLVSAHCTYDDRGFRYSPTQWTAFLGLHDSQSRAPGVQERLKRILSHPFNDFTF 708
Db 836 KMWLTVAHCF--EGR-----ENAAVWKVVGINNLDHPSV--PMQTRFVKTIILHPRYSRAVV 889
Qy 709 DYDIALLEKRPARYSSWVRPICLPDASHVFPACKALVNTGHWGTQYGGTGALLQSGEI 768

```

```

Db 890 DYDIALLEBEDISEFTGVAVPVLCLPNPEQWLEPPTYCYITGWH--NGNMKPFLLQSGEV 947
Qy 769 RVNIQTTCENLLPQO--ITPRMNCVGLSGVDSCGDSGGRPLSVEDGRIFGAGVYSWG 827
Db 948 RIISLEHCOSVFDKRTITTRMICAGYESGTVDSCMGDSGGLVCEKPGHWTLLFGLTSWG 1007
Qy 828 DGCAGR--NKPQVYRLLPLFRDWIK 850
Db 1008 SVCPSKVLGPGVYSNVSVFWEWIK 1031

RESULT 25
ID Q80YN4 PRELIMINARY; PRT; 1111 AA.
AC Q80YN4;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Corin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Sprague-Dawley;
RA Langenickel T., Pagel I., Buttgeriet J., Tenner K., Lindner M.,
RA Willebrock R., Dietz R., Bader M.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family 81.
DR EMBL, AY251285; AA086772.1; -.
DR HSBP, Q07954; ICR8.
DR GO, GO:0016020; C:membrane; IEA.
DR GO, GO:0004263; F:chymotrypsin activity; IEA.
DR GO, GO:0005044; F:scavenger receptor activity; IEA.
DR GO, GO:0004295; F:trypsin activity; IEA.
DR GO, GO:0007275; P:development; IEA.
DR GO, GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro, IPR000024; Fz_domain.
DR InterPro, IPR002172; LDL_receptor_A.
DR InterPro, IPR001254; Peptidase_S1.
DR InterPro, IPR001314; Peptidase_S1A.
DR InterPro, IPR009003; Pept_Set_Cys.
DR InterPro, IPR001190; Strc_Receptor.
DR Pfam, PF01392; Fz_2.
DR Pfam, PF00057; Ldl_recept_a; 6.
DR Pfam, PF00089; Trypsin_1.
DR PRINTS, PR00722; CHYMOTRYPSIN.
DR PRINTS, PR00261; LDLRECEPTOR.
DR SMART, SM00063; FRI; 2.
DR SMART, SM00192; LDLA; 7.
DR SMART, SM00202; SR; 1.
DR SMART, SM00202; TRYP_SPC; 1.
DR PROSITE, PS50038; Fz_2.
DR PROSITE, PS01203; LDLA_1_6.
DR PROSITE, PS50068; LDLA_2; 7.
DR PROSITE, PS50287; SRCR_2; 1.
DR PROSITE, PS50240; TRYPSIN_DOM; 1.
DR PROSITE, PS00135; TRYPSIN_SRR; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 1111 AA; 122635 MW; 3BA2706CBE81157F CRC64;

Query Match 14.8%; Score 694.5; DB 2; Length 1111;
Best Local Similarity 28.7%; Pred. No. 3,7e-39;
Matches 210; Conservative 80; Mismatches 259; Indels 183; Gaps 31;

Qy 205 TVQRTQNSGSGFLHARGVELMRTTGPFPDPSYPFAHARQOMALRGDADSVLSITFRSFD 264
Db 462 TFCQDGDGDCXGDSDEENCS-----DRPFP-----CPGDRGGLDS----- 497
Qy 265 LASGDE--RGSDDTVVNTL-----SPMEPHALVOLCGTYPPSYNLTFP----- 306

```

```

Db 498 --SCVESCASSSLCSDSSLSNCHCEPITL-ELCMNLP--YNLTHTYPNYIGHRTQKEAS 552
Qy 307 -SSQNVLLITLITTEHRHPFEATFPOLPRMSSCGRLKRAQGTFSPPYGHPP-NI 364
Db 553 ISMSALFPALVQNCYKYLMEFACITLIVPKCDVNTQ-----RVPCRL 597
Qy 365 DCTWNIIEVPNNQHVKVRPKFPYLLPEGVPACTCKDYVEINGEKYCGERSQFVTSNSNK 424
Db 598 LC-----EHSKRCESVL-----GIVGLQMPED-----TDCSQFPFQSSDNQ 634
Qy 425 IYRFHSDQSYDTGFLAEYLYSYDSSPDC-PCGTCTGTGRCIRKELACDGDADCTHSDSE 483
Db 635 TCL-----LPNEDVECSPSHFRCRSGRCVLSRRCDGQADCDSDSE 677
Qy 484 LNSCDAHQHPTCK-NKFKCPFLFWCVSDVDCGNSBOGCS-CPAOTFRCNSGKLSKS 541
Db 678 ENGCCRRLDMEGCLNNQCLKHTLICGFPCCSMBEKXCSFQDDELBECANECYPRD 737
Qy 542 QQCNKGKDDCGDSDPASCPRKXNVVTCYK-----TYRCLNG-----L 578
Db 738 LMCQGWTDSCSDSDMGC-----VTLKNGNSSFLTVHRSARDHVCADGMDLTSOLA 792
Qy 579 C--LSKNPE-----CDKE-----DSSGSDSK-----D 601
Db 793 CRQGLSEPVTYELVQCGEQQMLRLHSSWENLNGSTLQELVHRSCPSGSESLCTK 852
Qy 602 CDCGLRSFTR-QAVVGATDADEGEWPOVSLHALGQHICGSLISPNMLVSAHGYID 660
Db 853 QDCGRRAAAMNKRILIGRTSRPRMWMQCSLOSBSGHICGCLAKKWLTVAHCF-E 911
Qy 661 DRGRYDPTQWTAFLGLHDSQSRASGVQRRRLKRIISHPFENDFPDYDIALLEKRP 720
Db 912 GR----EDAVWKVYFVGINNDHPSG-FMOTRFVKTLLHPRYSRAVVDYDVSVELSD 966
Qy 721 AEXSWMRPCLPDASVFPAGKAIWYTGKGTGYGATLLOKGRIVNQTTCENL 780
Db 967 INETSYRVPCLSPREFLEBDTYCYITGMGH--MKNMPPKLOEGEVRILPLEQCCSYF 1024
Qy 781 POQ-ITPRMVCVGLSGVDSCQDSCGAPLSVYADRIFGAGVSWGDCAGR-NKPGV 838
Db 1025 DMKITTRMTCAGVESGTVDSCMGDSGPLVCERPGQWTLFGLTSGSVCFSLVLRPGV 1084
Qy 839 YTRLPLFRDWIK 850
Db 1085 YSNVSYFVDWIE 1096

```

```

RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Smi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsuno T., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hagiwara T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Horii F., Imclani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; AK038356; BAC29973.1; -.
DR MGD; MGI:1197523; P1887.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000858; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000998; MAM.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; Ldl_recept_a; 1.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00899; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.

```

```

DR PRINTS; PRO0020; MAMDOMAIN.
DR SMART; SM00042; CUB; 1.
DR SMART; SM00192; MAM; 1.
DR SMART; SM00137; MAM; 1.
DR SMART; SM00020; SR; 1.
DR SMART; SM00020; TRY Spec; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LDIRA_1; 1.
DR PROSITE; PS50068; LDIRA_2; 1.
DR PROSITE; PS00740; MAM_1; UNKNOWN_1.
DR PROSITE; PS50060; MAM_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Glycoprotein; Hydrolyase; Kinase; Protease; Serine protease.
PT NON TER
SQ SEQUENCE 777 AA; 87314 MW; A18E2F4ECF0ED3A8 CRC64;

Query Match 14.6% Score 684; DB 2; Length 777;
Best Local Similarity 29.3%; Pred. No. 1.3e-38;
Matches 186; Conservative 93; Mismatches 220; Indels 136; Gaps 22;

Qy 258 LTFRSFLASCDERGSGLVTVYNTLSPMEPHALVOLGCTPPSYNLTFHSSQNVLLITLI 317
Db 231 VVFFAFNRRGSGSTALADISTLNGICSGSPYEPETLVTPPP----- 272
Qy 318 TNRERRHPGFATFQLPNRRSCGG--RLKAQGTENSPYFGHYPPNIDCTNIEVFN 375
Db 273 -----ELP--TDCGPFELMEPNSTFSSPNPDYKYPQASCIWMLNQRG 315
Qy 376 QHVVRKFFFLLEPGVPAGCPDYVEI--NGE-----KYGERSQFVTSNSKIT 426
Db 316 KNIOHFOEFLN-----INDVVEVRDGGFSDLLAVYGPVKDLFTTNMT 367
Qy 427 VRFSDSYTDYGLAEYLIS---VDSDPG--PGQFTCTGRCIRKELKCDGMADCTHSD 482
Db 368 VIFTNMTETRRKRGKAPNTSGYIGIPEPCDDDFQCDGNCICILGNLCDSYPHCRGSD 427
Qy 483 ELNC-----SCDAGHQTCKNFKCPLFWCDSDVNDGSDNSDEQSCPAOTFRCSN 534
Db 428 EASCVRFLNGTRSNNGLVQFNHIS-----TWHI-----ACAEN-----WTOISN 467
Qy 535 GKCLSKSQOQNGKDDCCDGSDEACSP-----KYNVYTCIKHTYRCINGCLSKGN 584
Db 468 EVC-----HLLGIGSANSSMPISSTGGGFVRN-----QAPNGSLITPS 508
Qy 585 PECGKE---DCSDGDEKDCDGLRSFTROA--RVVGTDADEGEPMQVSLHALGQG 638
Db 509 LQGSQDILLQCNHKS-----CGEKVTOKVSPKIVGSDAQAGAMPVVALYHNDRS 562
Qy 639 H--ICGASLISPMVLVSAHCYIDNRGFRYSDEPTOWTAFGLHDGORSAPGVQERRLK 695
Db 563 TDRLLCGASLVSQMLVSAHCYVR---RMLDTRMTAVGLHQSMLTSPQVVRVRVD 618
Qy 696 RIIRHPFNDEFTFYDIALLELEKPAEYSSMVRICLPDASHVPRAGKALVWTGHTQY 755
Db 619 QIVINPHYDRRRKKNVDMHLEKRVNTDYOICLPSEEQOIFTPGRTCSIAWGVDKI 678
Qy 756 GGTGALLIKGEIVINQTCENTLPPQ--QTPRMKCVFLSGVDSCGSDGGLPSSVEA 814
Db 679 NGSTVDVPLKADVLSINKECQOOLPEVNTESMICAYEGEGGIDSCGSDGGLMCOE- 737
Qy 815 DGRIFGAGVSWGDCGGRNKPYYTTLPLFRMI 849
Db 738 NNRWFLVGVTSFGVOCALPNHGVYVRVSGTEMI 772

```

```

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
GN Name=PRSS7; Synonyms=ENTK;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Carnivora; Homnidae; Homo.
OC NCBI_TaxID=9606;
RX MEDLINE=21606074; PubMed=11719902;
RA Holzinger A., Maier E.M., Buck C., Mayerhofer P.U., Kappler M.,
RA Haworth J.C., Moroz S.P., Hadorn H.-B., Sadler J.E., Roscher A.A.;
RT "Mutations in the proenteropeptidase gene are the molecular cause of
RT congenital enteropeptidase deficiency.";
RL Am. J. Hum. Genet. 70:20-25(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=2089799; PubMed=10830953;
RA Hatori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Pak H.-S., Toyoda A., Ichii K., Tokoi Y., Choi D.-K., Groner Y.,
RA Soeda E., Onki M., Takagi T., Sakaki Y., Tauden S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudon J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Negamine K., Mitsuyama S., Antonakakis S.E.,
RA Mhoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Blocker H.,
RA Rammer J., Beck A., Klages S., Hennig S., Rieseemann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Leirach H., Reinhardt R., Yaspo M.-L.,
RL "The DNA sequence of human chromosome 21.";
RN Nature 405:311-319(2000).
RP [4]
RP SEQUENCE OF 749-1019 FROM N.A.
RX MEDLINE=94329561; PubMed=8052624;
RA Klatemco Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
CC -1- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys--|-Ile-7 bond in
CC trypsinogen.
CC -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC multimeric (heavy) chain linked by a disulfide bond.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (probable).
CC -1- TISSUE SPECIFICITY: Intestinal brush border.
CC -1- PFM: The chains are derived from a single precursor that is
CC cleaved by a trypsin-like protease.
CC -1- DISEASE: Defects in PRSS7 are a cause of enterokinase deficiency
CC (MIM:226200), a life-threatening intestinal malabsorption disorder
CC characterized by diarrhea and failure to thrive.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 MAM domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```







FT CARBOHYD 515 515 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 579 579 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 675 675 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 727 727 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 751 751 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 770 770 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 791 791 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 897 897 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 936 936 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 999 999 N-linked (G1CNAC. . .) (Potential).  
 SQ SEQUENCE 1069 AA; 118735 MW; E62549E46374C3D CRC64;  
 Query Match 14.4%; Score 673.5; DB 1; Length 1069;  
 Best Local Similarity 29.2%; Pred. No. 1e-37;  
 Matches 186; Conservative 93; Mismatches 220; Indels 137; Gaps 23;  
 QY 258 LTPSPFLASDEGSDLVVYVNLSPMEPHALVQLGYTPSPNLTFLHSSQVLLTLI 317  
 DB 522 VVENAFNRRGSGTALDDISLTNGICSGSPPEPTLPVTPPP----- 563  
 QY 318 TNERRRHGFEPATFPOLPRMSSCGG--RLRKAQGFSPSPYGHYPNIDCTWNIIEVNN 375  
 DB 564 -----ELP--TDGGPPELMEPNSTFSSPNPDKYPRQASCTWNLNLRG 606  
 QY 376 QHVKVRKFPYLLPEPVAGTCEPVYEI--NGS-----KYGERSQFVVTSSNKRIT 426  
 DB 607 KNIQLHFOEFDLEN-----INDVVEVRDGGFSDLSLLAVYTGPPVKDLFTSTNMT 658  
 QY 427 VRFHSDSYDTGRLAYLS---YDSSDPC--PGQFTGRTGRCIKELRCQGMACTHSD 482  
 DB 659 VIFTNMETRRKGRKANFTSGYVIGIEPCDDDFQCKDGCIPVLGNLCOSYPHCRGSD 718  
 QY 483 ELNC-----SCDAHQFTCKKFKCKPLFWVCDSDVNDGSDNGSCSPAQTFRCSN 534  
 DB 719 EASGVRLNGRNSNGLVQFNHIS-----IMH-----ACAE-----WTQISN 758  
 QY 535 GKCLSKSQCKNGKDCDGDSEACSP-----KVVVYTCYKHTYCLNGCLSKGN 584  
 DB 759 EVC-----HLLGLGSANSSMPISSTGCGPFAVYN-----QAPNGSLITPS 799  
 QY 585 PECDGKE-----DCSDGDEKDCDGLASFTRQ--RVVGGDADDEGMPQVSLHALGCG 638  
 DB 800 LQCSQDSLILLCNHS-----CGEKVKYQKVPKIVGSGDAQAQAMPVVALYHRDS 853  
 QY 639 H---ICGASHISPMVLAAHCYIDRGRFSDPTQWTAFLGLIDGQSRAPQVGERRLK 695  
 DB 854 TDRLLCASLVSSMVLVSAHCVR-----RNLDPTRMTAVGLHMGSLNSPQVVRVVD 909  
 QY 696 RIISHPFNDFTFDYDIALLEKPAVSSMVRDICTLDASHVFPACKAIVWTGHTQY 755  
 DB 910 QIVINPHYDRRKRVNDIAMHMLEKRVNVTYDICTLPEEQIFIPORTCSIAQMGYDKI 969  
 QY 756 -GGGALLQKGEIRVNOTTCENLLPQ-QTPPRMVGVSFGSGVSDCGDSSGSLSSVE 813  
 DB 970 NAGSTVDVLEADVPLISNEKCOQDLPEYNTSMTICAGYBEGGIDSCQDGSGLPMLCOE 1029  
 QY 814 ADGRIFGAGVSWDGCAGRNKPGVYTRLPFLFRMI 849  
 DB 1030 -NNRMFLVGVTSFVGQCALPHHGPVYRVSGFIEMI 1064  
 RESULT 29  
 CORI MOUSE STANDARD; PRT; 1113 AA.  
 AC 092319:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-  
 DE converting enzyme) (Corin) (Low density lipoprotein receptor related  
 DE protein 4).  
 GN Name=Corin; Synonyms=Crm, Lrp4;  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98429596; PubMed=9756624;  
 RA Tomita Y., Kim D.-H., Maqsoori K., Fujino T., Yamamoto T.T.;  
 RT "A novel low-density lipoprotein receptor-related protein with type II  
 RT membrane protein-like structure is abundant in heart."  
 RL J. Biochem. 124:784-789(1998).  
 CC -1- FUNCTION: Converts pro-ANP to ANP. Cleaves pro-ANP specifically  
 CC between Arg-122 and Ser-123 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in heart.  
 CC -1- SIMILARITY: Belongs to peptidase family 31.  
 CC -1- SIMILARITY: Contains 2 frizzled (Fz) domains.  
 CC -1- SIMILARITY: Contains 7 LDL-receptor class A domains.  
 CC -1- SIMILARITY: Contains 1 SRCR domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB013874; BAA34371.1; -.  
 DR PIR; JEO315; JEO315.  
 DR HSSP; P01130; 1A3T.  
 DR MEROPS; S01.019; -.  
 DR MGD; MGI.1349451; Corin.  
 DR InterPro; IPR000024; Fz domain.  
 DR InterPro; IPR002172; LDL\_receptor\_A.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR InterPro; IPR009003; Pept\_Set\_Cys.  
 DR InterPro; IPR001190; Srcr\_receptor.  
 DR Pfam; PF01392; Fz; 2.  
 DR Pfam; PF00057; Ldl\_recept\_a; 6.  
 DR Pfam; PF00089; Trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR SMART; SM00063; FRI; 2.  
 DR SMART; SM00192; LDLA; 7.  
 DR SMART; SM00202; SR; 1.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PSS0038; Fz; 2.  
 DR PROSITE; PSS01209; LDLRA\_1; 6.  
 DR PROSITE; PSS0068; LDLRA\_2; 7.  
 DR PROSITE; PSS00420; SRCR\_1; FALSE\_NEG.  
 DR PROSITE; PSS0287; TRCP\_2; 1.  
 DR PROSITE; PSS0240; TRYPsin DOM; 1.  
 DR PROSITE; PSS00134; TRYPsin\_HIS; FALSE\_NEG.  
 DR PROSITE; PSS00135; TRYPsin\_SER; 1.  
 KM Glycoprotein; Hydroxylase; Repeat; Serine protease; Signal-anchor;  
 KM Transmembrane.  
 FT DOMAIN 1 112  
 FT TRANSMEM 113 133  
 FT DOMAIN 134 1113  
 FT DOMAIN 201 327  
 FT DOMAIN 336 372  
 FT DOMAIN 373 408  
 FT DOMAIN 409 445  
 FT DOMAIN 446 483  
 FT DOMAIN 518 641  
 FT DOMAIN 647 682  
 FT DOMAIN 683 721  
 FT DOMAIN 722 757  
 FT DOMAIN 758 853  
 FT DOMAIN 869 1113  
 FT ACT\_SITE 910 910  
 Cytoplasmic (Potential).  
 Signal-anchor for type II membrane  
 protein (Potential).  
 Extracellular (Potential).  
 Fz 1.  
 LDL-receptor class A 1.  
 LDL-receptor class A 2.  
 LDL-receptor class A 3.  
 LDL-receptor class A 4.  
 Fz 2.  
 LDL-receptor class A 5.  
 LDL-receptor class A 6.  
 LDL-receptor class A 7.  
 SRCR.  
 Serine protease.  
 Charge relay system (By similarity).

```

FT ACT SITE 959 959 Charge relay system (By similarity).
FT ACT_SITE 1052 1052 Charge relay system (By similarity).
FT DISULFID 857 979 By similarity.
FT DISULFID 895 911 By similarity.
FT DISULFID 1022 1037 By similarity.
FT DISULFID 1048 1077 By similarity.
FT CARBOHYD 147 147 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 202 202 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 208 208 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 298 298 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 317 317 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 373 373 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 411 411 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 444 444 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 481 481 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 519 519 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 537 537 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 635 635 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 719 719 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 765 765 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 828 828 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 970 970 N-linked (G1cNAc . .) (Potential).
FT CARBOHYD 1089 1089 N-linked (G1cNAc . .) (Potential).
SQ SEQUENCE 1113 AA; 122984 MW; B845B2C5F20DD8EC CRC64;

Query March 14.1%; Score 661.5; DB 1; Length 1113;
Best Local Similarity 32.7%; Pred. No. 7.1e-37;
Matches 158; Conservative 67; Mismatches 187; Indels 71; Gaps 16;

Qy 428 RFDSDSYDTDTGFLAEYLSDSPDPCQFTCRGCRKELRCGMDCTDHSDELNC 487
Db 627 QFPRESSDNDQCLP---NEDVECSPEHFKCRGRCLSRRCGDQADCDDEBENCG 683
Qy 488 CDAGHQFTCK-NKFCPLFWVCDVNDGNSDEQGS-CPAQFRCSNGKCLSKSQCN 545
Db 684 CKEBALMECPFNKCKLKTLLICDGFPPDCPSMDEKNGCSFCODNLELCANHECVPRDLMD 743
Qy 546 GKDDCGSGSDASCPKNV-----VVTCTKHTYCLNG-----LC----- 579
Db 744 GWVDCSSSDDEWGCVTLSKNGNSSLLTVHKSAAEHVCAQWRETTLSQLACKOMGLGEP 803
Qy 580 -LSKGNPECDGKE-----DCSDGSDKED-----C--DCGLNSFT 610
Db 804 SVTLITIGQEGQQLRLVPMNENLNGSTLQELVYRHSCPRSRSLSCSDQDGRRAA 863
Qy 611 R-QARVVGTDADGEPMQVSLHALGQGHICGSLISPNVLVSANHCYIDRGRISDP 669
Db 864 RMNKRILIGRTSRGRPMWQCSLQSEPSGHICGCVLLAKKWLIVAHCF-EGR----EDA 918
Qy 670 TQWNAFLGHLHQSGRSARGVVERLRKLIHPFNDFPDYDIALLEKRAEYSSWVR 729
Db 919 DVMKRVFGINLLDHPG-FMQTRFVKTLIHPRSRAVVDYDISVELSDINETSVP 977
Qy 730 ICLPDASHVFAKAIWVTGNGHTOYGGTGALLLOKGRIRVYNOTGENTLPPQ-ITPRM 788
Db 978 VCLSPREYLEPDYCYITGNGH--MKNMFPKIQEGSVRLIPLBQCSYIDMKITTKM 1035
Qy 789 MCVFPLSGVDSCGDSGGLPLSSVEADGRIFGAGVNSGDCAGR-NKPGVYTLPLFRD 847
Db 1036 ICAGYEGTVDSCMGDSGGLVRCERPGQWTLFGLTSGVSCFSLVGLPGVYSNVSYVG 1095
Qy 848 MIR 850
Db 1096 WIE 1098

RESULT 30
Q9DGR2 PRELIMINARY; PRT; 767 AA.
AC Q9DGR2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

```

```

DE Embryonic serine protease-2.
GN Name=Xesp-2;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=20363741; Pubmed=10903452;
RA Yamada K., Takahatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
from Xenopus laevis."
RL Gene 252:209-216 (2000).
-1- SIMILARITY: Belongs to peptidase family S1.
EMBL; AB038497; BAB08217.1; -.
DR HSSP; P00760; IEZX.
DR MEROPS; S01.049; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 8.
DR SMART; SM00020; Tryp_spc; 1.
DR PROSITE; PS01209; LDLRA_1; 8.
DR PROSITE; PS01068; LDLRA_2; 2.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrobase; Protease; Serine protease.
SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE96E CRC64;

Query March 13.7%; Score 639.5; DB 2; Length 767;
Best Local Similarity 26.6%; Pred. No. 1.5e-35;
Matches 170; Conservative 71; Mismatches 215; Indels 183; Gaps 17;

Qy 365 DCTWN-----IEVNNQHVYKRFKPYLLERGVAGTCPKDYVEINGERKYC--GERSQ 415
Db 150 DCPYGDDEBRCATPTPTPCQWYCSYV-----TCYGVQICNGVQDCPYGDDE 200
Qy 416 FVYTSNKNKTVRPHSDSYDTGFLAEYLSDSPDPCQFTCRG----- 462
Db 201 NCATKTPSIFTCQWYC--SYTSTCIYGVQI-----CNGVQDCAYDDEBRCATKTPSI 251
Qy 463 -----RCIKRELRCDGMDCTDHSDELNCSDAHQFTCK-----NKRCKP 503
Db 252 PTCQLYSYYTYTCLYAVQICNGVLDCPFVDERKCVIATSTPCQIYCKDPMFYTCIY 311
Qy 504 LFWVCDVNDGNSDDEQGC-----SCPAQFRCSN-----GKCLSKSQCNCKD 548
Db 312 AYQWCDGVRCYQYDDERNCVATATTTATTTSPPTCOIYCMNFRYYTYTCLYAVQMGQVR 371
Qy 549 DCGDGSDEBASC-PKVNVTCTKHTYRCLNGCLSKGNPECDGKEDGSDGSDKDC----- 602
Db 372 QCYYGDDELNCDFRTTAYCEK--RCGSSSVSCVLSQWCDGVSDCPYGBDEMSCVSLYP 428
Qy 603 ----- 602
Db 429 ADFOLOVYTSVSAMLPVCSDDYMNDFGRPAQDPFGNGSSVNRYYTLMSPIYAPNGYFKL 488
Qy 603 -----DCGLRSFTROARVVGTDADGEPMQVSL 632

```

```

Db      489 YSGYMRSKFYTSVQYSSYCYSGNVVSLHCISCGVSNNSLVSRIVGTFANLGNMPWQVNL 548
Qy      633 HALQOGHICGASLISPMNLVSAAHCIYIDRGFRYSDDPTQWTAFLGLHDOSORSAFVQER 692
Db      549 QYI-TGYLCGSIISPKMIVTAHCVYGS---YSSASGMRVPAG--TLTKPSYTNASAY 601
Qy      693 RLKRIISHPPFNDFTFDYIALLELEKPAEYSSMVRPICLPDASHVPAGKAIWVTGWH 752
Db      602 FVERIIVHPGYSXYTNDIALMKLRDEITFGYTTQPVCLPNSGMFWBAGTTTWISGWS 661
Qy      753 TVYGTGALLIÖKEIRVINOTTCEN--LBPQITPRMVCVFLSGGVDSCQDSCGGLS 810
Db      662 TYEGGSVSTYLYQYALPLIDSNVCNOSYVYNGQITSSMICAGYLSGVDTCCQDSCGGLV 721
Qy      811 SVEADGRIFGAGVYVSWGDCAGRNKPGVYTRLPFLFRDMI 849
Db      722 N-KRNGTWMVLVGDTSKWDGCAKANKPGVYGNVTTFLEMI 759

```

Search completed: November 29, 2004, 08:32:31  
 Job time : 150.315 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:19:13 ; Search time 113.685 Seconds

(without alignments)  
3456.743 Million cell updates/sec

Title: US-09-936-333-5

Perfect score: 3789

Sequence: 1 MAERVYMLPPRRASLSKFV.....PGVYTRLPFRMDIKENTGV 683

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 100 summaries

Database : 1: uniprot\_02.\*

2: uniprot\_sprot.\*  
3: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3760	99.2	855	1	ST14_HUMAN
2	3209	84.7	855	1	SP14_MOUSE
3	3209	84.7	855	2	BAC35125
4	3197	84.4	855	2	Q9JUC17
5	2379	62.8	422	2	Q8WVC1
6	2280.5	60.2	845	2	Q6GR54
7	2254.5	59.5	845	2	Q9DGR1
8	1543	40.7	663	2	Q6DEV0
9	1017.5	26.9	572	2	Q7RTY8
10	1017.5	26.9	572	2	BAD18401
11	1015.5	26.8	572	2	Q8BIR6
12	1014	26.8	799	2	Q6PF94
13	1014	26.8	799	2	AAH57674
14	1014	26.8	811	1	TMS6_MOUSE
15	1007	26.6	811	1	TMS6_HUMAN
16	1005	26.5	802	2	Q6UXD8
17	1005	26.5	802	2	AAQ88764
18	985	26.0	824	2	Q6ICC2
19	985	26.0	824	2	CAG30332
20	718.5	19.0	1034	1	ENTK_PIG
21	717.5	18.9	855	2	Q7Z410
22	717.5	18.9	1059	2	Q7Z411
23	703.5	18.6	1035	1	ENTK_BOVIN
24	694	18.3	1042	1	CORI_HUMAN
25	690.5	18.2	1111	2	Q80YN4
26	688	18.0	1019	1	Q8CANG
27	683.5	18.0	1019	1	ENTK_HUMAN
28	677.5	17.9	1069	1	ENTK_MOUSE
29	663.5	17.5	1113	1	CORI_MOUSE
30	644.5	17.0	767	2	Q9DGR2
31	644	17.0	722	2	Q6NUPS

32	644	17.0	722	2	AAH68636	AAH68636 xenopus 1
33	632.5	16.7	680	2	Q868H7	Q868H7 branchiocto
34	622	16.4	581	2	Q9XZM7	Q9XZM7 strongyloce
35	619.5	16.3	680	2	Q868H5	Q868H5 branchiocto
36	608	16.0	688	2	Q868H6	Q868H6 branchiocto
37	601	15.9	490	2	Q7TND4	Q7TND4 mus musculu
38	600	15.8	490	1	TMS2_MOUSE	Q9J48 mus musculu
39	598.5	15.8	490	2	Q6P7D7	Q6P7D7 rattus norv
40	598.5	15.8	490	2	AAH61712	AAH61712 rattus no
41	598	15.8	490	2	Q920K3	Q920K3 rattus norv
42	593	15.7	688	2	Q868H4	Q868H4 branchiocto
43	588	15.5	422	1	DEB1_HUMAN	Q9UL2 homo sapien
44	588	15.5	423	2	Q6UW31	Q6UW31 homo sapien
45	588	15.5	423	2	AAQ89376	AAQ89376 homo sapi
46	584.5	15.4	701	2	Q9J3S9	Q9J3S9 rattus norv
47	584.5	15.4	868	2	Q9Y1V3	Q9Y1V3 polyandroca
48	581	15.3	703	2	Q7RTY7	Q7RTY7 homo sapien
49	578	15.3	704	1	Q8CHN8	Q8CHN8 rattus norv
50	577.5	15.2	704	1	CRAR_MOUSE	P98064 mus musculu
51	576.5	15.2	1524	2	Q91674	Q91674 xenopus lae
52	571	15.1	676	2	Q6DUI5	Q6DUI5 rattus norv
53	571	15.1	676	2	Q6DUI6	Q6DUI6 cyprius ca
54	565	14.9	699	1	CRAR_HUMAN	P48740 h complemen
55	558.5	14.7	492	1	TMS2_HUMAN	Q15393 homo sapien
56	558.5	14.7	492	2	Q96T73	Q96T73 homo sapien
57	558.5	14.7	492	2	AAH51839	AAH51839 homo sapi
58	557.5	14.7	423	2	Q8BM10	Q8BM10 mus musculu
59	557	14.7	681	2	Q7ZT70	Q7ZT70 lampetra ja
60	553	14.6	698	2	Q6GPE9	Q6GPE9 xenopus lae
61	550	14.5	307	2	Q6ZND6	Q6ZND6 homo sapien
62	550	14.5	307	2	BAD18439	BAD18439 homo sapi
63	549.5	14.5	638	2	Q8R0P5	Q8R0P5 mus musculu
64	547	14.4	453	2	Q812A6	Q812A6 mus musculu
65	546	14.4	453	1	TMS3_MOUSE	Q812A6 mus musculu
66	545.5	14.4	638	1	KAL_MOUSE	P26262 mus musculu
67	543.5	14.3	453	2	Q6ZMC3	Q6ZMC3 homo sapien
68	543.5	14.3	453	2	AAQ88823	AAQ88823 homo sapi
69	543.5	14.3	453	2	BAD18806	BAD18806 homo sapi
70	541.5	14.3	417	2	Q8VH14	Q8VH14 rattus norv
71	539	14.2	707	2	Q8QGV0	Q8QGV0 cyprius ca
72	535.5	14.1	279	2	Q9QZ74	Q9QZ74 rattus norv
73	533	14.1	454	1	TMS3_HUMAN	P57727 homo sapien
74	528.5	13.9	617	2	Q8JIS1	Q8JIS1 triakis scy
75	526.5	13.9	618	1	KAL_RAT	P14222 rattus norv
76	525.5	13.9	417	2	Q8VDY1	Q8VDY1 mus musculu
77	525.5	13.9	417	2	Q8VHX8	Q8VHX8 mus musculu
78	523.5	13.8	279	2	Q7TNX3	Q7ZT2 xenopus lae
79	521	13.8	600	2	Q7ZTR2	Q61139 mus musculu
80	518	13.7	603	1	CPAI_MOUSE	Q9WUJ3 rattus norv
81	518	13.7	604	2	CPAI_RAT	Q9WU1 xenopus lae
82	515.5	13.6	698	2	Q9PU71	Q9PU71 homo sapien
83	513.5	13.6	537	2	Q9BYE1	Q9BYE1 homo sapien
84	513.5	13.6	581	2	Q9BYE2	Q9BYE2 mus musculu
85	513	13.5	455	1	TMS5_MOUSE	Q9E04 mus musculu
86	513	13.5	471	2	Q8CFE0	Q8CFE0 mus musculu
87	512	13.5	666	2	Q6O1O9	Q6C19 gallus gall
88	512	13.5	686	2	AAH73179	AAH73179 gallus ga
89	511.5	13.5	418	1	HATT_HUMAN	Q6U235 homo sapien
90	510	13.5	645	2	Q7PWE4	Q7PWE4 anopheles g
91	508.5	13.4	445	2	Q8CJ17	Q8CJ17 rattus norv
92	508.5	13.4	467	2	Q967X8	Q967X8 panulirus a
93	507.5	13.4	250	2	Q9V514	Q9V514 drosophila
94	507.5	13.4	438	2	Q6ZWK6	Q6ZWK6 homo sapien
95	507.5	13.4	438	2	BAC85495	BAC85495 homo sapi
96	506.5	13.4	455	2	Q8CDR0	Q8CDR0 mus musculu
97	506	13.4	613	2	Q03711	Q03711 xenopus lae
98	506	13.4	719	2	Q6DJ90	Q6DJ90 xenopus tro
99	504.5	13.3	558	2	Q86YH4	Q86YH4 homo sapien
100	504	13.3	790	1	PLAN_PIG	P06867 sus scrofa

## ALIGNMENTS

RESULT 1  
ST14\_HUMAN STANDARD; PRT; 855 AA.  
ID Q9Y5T6; Q9B501; Q9H3S0; Q9H3B6; Q9HCA3;  
AC 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matrilysin) (Membrane-type serine protease 1) (MT-SPL) (Prostasin) (Serine protease TAD6-15) (Tumor associated differentially-expressed gene-15 protein).  
GN Name=ST14; Synonyms=PRSS14, SNCL19, TADG15;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99303581; PubMed=10373424;  
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;  
RT "Molecular cloning of cDNA for matrilysin, a matrix-degrading serine  
RT protease with trypsin-like activity";  
RL J. Biol. Chem. 274:118231-118236 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99432178; PubMed=10500122;  
RA Takeuchi T., Shuman M.A., Craik C.S.;  
RT "Reverse biochemistry: use of macromolecular protease inhibitors to  
RT dissect complex biological processes and identify a membrane-type  
RT serine protease in epithelial cancer and normal tissue";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061 (1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Prostate;  
RA Yamaguchi N., Mitsui S.;  
RT "Molecular cloning of a novel transmembrane serine protease expressed  
RT in human prostate";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,  
RA O'Brien T.J.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Blood, and Muscle;  
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein W.J., Umed T.B., Tomihataki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmitt J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [6]  
RP SEQUENCE OF 340-664 FROM N.A.  
RA Cao J., Fan W., Zheng S.;  
RT "Genomic analysis of a novel human serine protease SNCL9";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [7]

RP CHARACTERIZATION.  
RX TISSUE=Milk;  
RC MEDLINE=99303582; PubMed=10373425;  
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;  
RT "Purification and characterization of a complex containing matrilysin  
RT and a Kunitz-type serine protease inhibitor from human milk";  
RL J. Biol. Chem. 274:18237-18242 (1999).  
CC -1- FUNCTION: Degrades extracellular matrix. Proposed to play a role  
CC in breast cancer invasion and metastasis. Exhibits trypsin-like  
CC activity as defined by cleavage of synthetic substrates with Arg  
CC or Lys as the P1 site.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
CC -1- SIMILARITY: Belongs to peptidase family 51.  
CC -1- SIMILARITY: Contains 2 CUB domains.  
CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF118224; AAD42765.2; -;  
DR EMBL; AF133086; AAF00109.1; -;  
DR EMBL; AB030036; BAB20376.1; -;  
DR EMBL; AF057145; AAG15395.1; -;  
DR EMBL; BC005826; AAH05826.2; -;  
DR EMBL; BC030532; AAH0532.1; -;  
DR EMBL; AF283256; AAG13949.1; -;  
DR PDB; 1EAW; X-ray; A/C=615-855.  
DR MEROPS; S01.302; -;  
DR Genew; HGNC:11344; ST14.  
DR MIM; 606797; -;  
DR GO; GO:0008687; C:integral to plasma membrane; TAS.  
DR GO; GO:0008236; F:serine-type peptidase activity; TAS.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.  
DR InterPro; IPR000859; CUB.  
DR InterPro; IPR002172; LDL\_receptor\_A.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR Pfam; PF00431; CUB; 2.  
DR Pfam; PF00057; Ldl\_recept\_a; 4.  
DR Pfam; PF00089; Trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00261; LDLRECEPTOR.  
DR SMART; SM00192; LDLa; 4.  
DR SMART; SM00020; TRYP\_SPC; 1.  
DR PROSITE; PS01180; CUB; 2.  
DR PROSITE; PS01209; LDLRA\_1; 2.  
DR PROSITE; PS0068; LDLRA\_2; 4.  
DR PROSITE; PS00240; TRYP\_SIN\_DOM; 1.  
DR PROSITE; PS00134; TRYP\_SIN\_HIS; 1.  
DR PROSITE; PS00135; TRYP\_SIN\_SER; 1.  
KW 3D-structure; Glycoprotein; Hydrolase; Repeat; Serine protease;  
KW Signal-anchor; Transmembrane.  
FT DOMAIN 1 55 Cytoplasmic (Potential).  
FT TRANSMEM 1 76 Signal-anchor for type II membrane  
FT protein (Potential).  
FT DOMAIN 77 855 Extracellular (Potential).  
FT DOMAIN 214 334 CUB 1.  
FT DOMAIN 340 447 CUB 2.  
FT DOMAIN 452 487 LDL-receptor class A 1.  
FT DOMAIN 487 524 LDL-receptor class A 2.  
FT DOMAIN 524 560 LDL-receptor class A 3.  
FT DOMAIN 566 603 LDL-receptor class A 4.  
FT DOMAIN 615 854 Serine protease.  
FT ACT\_SITE 656 Charge relay system (by similarity).  
FT ACT\_SITE 711 Charge relay system (by similarity).  
FT ACT\_SITE 805 Charge relay system (by similarity).  
FT ACT\_SITE 805

FT CARBOHYD 109 109 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 302 302 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 485 485 N-linked (G1CNAC. . .) (Potential).  
 FT CARBOHYD 772 772 N-linked (G1CNAC. . .) (Potential).  
 FT CONFLICT 381 381 A -> S (in Ref. 4).  
 FT CONFLICT 674 674 R -> S (in Ref. 3).  
 SQ SEQUENCE 855 AA; 94769 MM; 2614313201F99C9 CRC64;

Query Match 99.2%; Score 3760; DB 1; Length 855;  
 Best Local Similarity 99.4%; Pred. No. 3.8e-236;  
 Matches 679; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRASLSKSFVTVSVAPPTDSTQRTQDNCSFGHARGLVLRFTTPG 60  
 DB 173 MAERVVMLPPRASLSKSFVTVSVAPPTDSTQRTQDNCSFGHARGLVLRFTTPG 232  
 QY 61 FPDSPYPAHARQNALRGDADSVLSLTFRSPDLASCDERGSGLVTVNTLSPMEPHALVQ 120  
 DB 233 FPDSPYPAHARQNALRGDADSVLSLTFRSPDLASCDERGSGLVTVNTLSPMEPHALVQ 292  
 QY 121 LCGTTPSYNLTFRSSQNVLLITLTNTERHHPGEAFEPOLPRMSSCGGLRKAQGTFN 180  
 DB 293 LCGTTPSYNLTFRSSQNVLLITLTNTERHHPGEAFEPOLPRMSSCGGLRKAQGTFN 352  
 QY 181 SPYRGHYPPNIDCTMNTEVNNQHVKKRFFYLLEPRACGCPKDYVEINGEKCGE 240  
 DB 353 SPYRGHYPPNIDCTMNTEVNNQHVKKRFFYLLEPRACGCPKDYVEINGEKCGE 412  
 QY 241 RSQFVTVSNKNTIVRFHSDSYTDGFLAEYLSYSSDSPCGQFTCRGRCIRKELRCD 300  
 DB 413 RSQFVTVSNKNTIVRFHSDSYTDGFLAEYLSYSSDSPCGQFTCRGRCIRKELRCD 472  
 QY 301 GMADCTHSDSLNCSGAGHFTCKNCKRCLFWVCDSVNDCCGNSDPOGSCCAQGTFR 360  
 DB 473 GMADCTHSDSLNCSGAGHFTCKNCKRCLFWVCDSVNDCCGNSDPOGSCCAQGTFR 532  
 QY 361 SNGKCLSKSQCGKDGCDGSDDEASCPKVVVCTKRTYCLNGTCLSKGNPECDGKED 420  
 DB 533 SNGKCLSKSQCGKDGCDGSDDEASCPKVVVCTKRTYCLNGTCLSKGNPECDGKED 592  
 QY 421 CSDSDKDCDCCGLRSFTQARVVGTDADGEMPMWQSLHALQGHICGASLISPMWLV 480  
 DB 593 CSDSDKDCDCCGLRSFTQARVVGTDADGEMPMWQSLHALQGHICGASLISPMWLV 652  
 QY 481 SAACHYIDDRGFRSDPTQMTAFGLHDSQSRAPGVQERLKLIIHPFNDDTFDYDI 540  
 DB 653 SAACHYIDDRGFRSDPTQMTAFGLHDSQSRAPGVQERLKLIIHPFNDDTFDYDI 712  
 QY 541 ALLELEKPAEYSSWVRPCLBDASHVPAPKAIWVGHTQYGGTGLIIQKEIRVYN 600  
 DB 713 ALLELEKPAEYSSWVRPCLBDASHVPAPKAIWVGHTQYGGTGLIIQKEIRVYN 772  
 QY 601 QTTCENTLLPQOITPRMPCVGLSGGVDSQCGDSGGPLSSVEADRIFOAGVSVMGDCAQ 660  
 DB 773 QTTCENTLLPQOITPRMPCVGLSGGVDSQCGDSGGPLSSVEADRIFOAGVSVMGDCAQ 832  
 QY 661 RNKPGVYTRLPFRDMIKENTGV 683  
 DB 833 RNKPGVYTRLPFRDMIKENTGV 855

RESULT 2  
 ST14 MOUSE STANDARD, PRT, 855 AA.  
 AC P56677;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-JUN-2004 (Rel. 44, Last annotation update)  
 DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Epithin).  
 GN Name=St14; Synonym=Prs14;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;  
 RX MEDLINE=99216440; PubMed=10199918;  
 RA Kim M.G., Chen C., Lyy M.S., Cho E.G., Park D., Kozak C.,  
 RA Schwartz R.H.;  
 RT "Cloning and chromosomal mapping of a gene isolated from thymic  
 RT stromal cells encoding a new mouse type II membrane serine protease,  
 RT epithin, containing four LDL receptor modules and two CUB domains.";  
 RT Immunogenetics 49:420-428(1999).  
 RL [2]  
 RN REVISIONS TO 23, 321, 325, 343; 409-410 AND C-TERMINUS.  
 RP STRAIN=C.B.17SCID; TISSUE=Thymus;  
 RC Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Vallatton D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bonfield G.G.,  
 RA Blakeley K.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalka U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -1- TISSUE SPECIFICITY: Highly expressed in intestine, kidney, lung,  
 CC and thymus. Not expressed in skeletal muscle, liver, heart, testis  
 CC and brain.  
 CC -1- SIMILARITY: Belongs to peptidase family S1.  
 CC -1- SIMILARITY: Contains 2 CUB domains.  
 CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)  
 CC or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).

CC EMBL, AF042822; AA002230.3; -;  
 CC EMBL, BC005496; AA005496.1; -;  
 CC HSSP, P00760; IEZX.  
 CC MEROPS, S01.302; -;  
 CC MGD, MGI:1338881; St14.  
 CC GO, GO:0005576; C:extracellular; IDA.  
 CC GO, GO:0019897; C:extrinsic to plasma membrane; IDA.  
 CC GO, GO:0008236; F:serine-type peptidase activity; IDA.  
 CC InterPro, IPR000859; CUB.  
 CC InterPro, IPR002172; LDL\_receptor\_A.  
 CC InterPro, IPR001254; Peptidase\_S1.  
 CC InterPro, IPR001314; Peptidase\_S1A.  
 CC InterPro, IPR009003; Pept\_Ser\_Cys.  
 CC Pfam, PF00431; CUB; 2.  
 CC Pfam, PF00057; Ldl\_recept\_a; 4.  
 CC Pfam, PF00089; Trypsin\_1.  
 CC PRINTS, PR00722; CHYMOTRYPSIN.

```

DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLA; 4.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLA_1; 2.
DR PROSITE; PS50068; LDLA_2; 4.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Glycoprotein; Hydroxylase; Repeat; Serine protease; Signal-anchor;
KM Transmembrane.
KT DOMAIN 1 55 Cytoplasmic (Potential).
FT TRANSMEM 56 76 Signal-anchor for type II membrane
FT protein (Potential).
FT FT Extracellular (Potential).
FT DOMAIN 77 855 CUB 1.
FT DOMAIN 214 331 CUB 2.
FT DOMAIN 340 444 LDL-receptor class A 1.
FT DOMAIN 451 488 LDL-receptor class A 2.
FT DOMAIN 489 522 LDL-receptor class A 3.
FT DOMAIN 523 561 LDL-receptor class A 4.
FT DOMAIN 565 604 Serine protease.
FT DOMAIN 615 854 Charge relay system (By similarity).
FT ACT_SITE 656 656 Charge relay system (By similarity).
FT ACT_SITE 711 711 Charge relay system (By similarity).
FT ACT_SITE 805 805 Charge relay system (By similarity).
FT CARBOHYD 107 107 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 302 302 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 365 365 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 421 421 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 489 489 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 772 772 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 855 AA; 94654 MW; 4F10B84DA2146DD5 CRC64;

Query Match 84.7%; Score 3209; DB 1; Length 855;
Best Local Similarity 83.2%; Pred. No. 2.3e-217;
Matches 568; Conservative 59; Mismatches 56; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRARSLKFSVTVSVAPPTDSKTVQRTQDNSSCGFLHARGVELARFTTPG 60
Db 173 MAVRVVTLPPRARALSLFVLTSTVAFPIDPRMLQRTQDNSSCGFLHARGVELARFTTPG 232
Qy 61 PPDSPYAAHARQWALRGDADSVLSLTFRSPDLASCDERGSGLVTVYNTLSMPMHALVQ 120
Db 233 FPNSPYAAHARQWALRGDADSVLSLTFRSPDLASCDERGSGLVTVYNTLSMPMHALVQ 292
Qy 121 LCGFPPSYNTLTHSSQNVLLTLITTEPRHHPGFATFQLPMSGCGRLRAQCTFN 180
Db 293 LCGFSPSYNTLTHSSQNVLLTLITTEPRHHPGFATFQLPMSGCGRLRAQCTFN 352
Qy 181 SPYYPGHYPNIDCTWNIIEVNNQHVYKFKFYLLPRRACGTCPPDQYVEINKEKCGE 240
Db 353 SPYYPGHYPNIDCTWNIIEVNNQHVYKFKFYLLPRRACGTCPPDQYVEINKEKCGE 412
Qy 241 RSQFVTSNSNKTIVRPHSDQSYTDTGFLAEVLSYDSSDPGQFTCRGCIKRLKELCD 300
Db 413 RSQFVTSNSNKTIVRPHSDQSYTDTGFLAEVLSYDSSDPGQFTCRGCIKRLKELCD 472
Qy 301 GMADCTHSDPELNCSDAGHGFQCKNKFCKPLFVNCVSVNDGNSNBOGCSGCAQFRC 360
Db 473 GMADCTHSDPELNCSDAGHGFQCKNKFCKPLFVNCVSVNDGNSNBOGCSGCAQFRC 532
Qy 361 SNGKCLSKSQOQKNGKDCGDSGDSASCPKVVVVTCTGHTVRCLNGLCLSKNGPECDEKED 420
Db 533 SNGKCLPQSQKCKNCKDKGDSGDSASCDVNVVVSCTITTRCQNGLCLSKNGPECDEKED 592
Qy 421 CSQSGDEKDCDCGIRSFTRQARVVGGTADGEMPMQVSLHALGQGHICGASLISPMVLY 480
Db 593 CSQSGDEKDCDCGIRSFTRQARVVGGTADGEMPMQVSLHALGQGHICGASLISPMVLY 652
Qy 481 SAACHYIDDDGFRSDPTQMTAFGLHDSQSRAPGVQERIKRIISHFPNDPTFYDI 540
Db 653 SAACHYIDDDGFRSDPTQMTAFGLHDSQSRAPGVQERIKRIISHFPNDPTFYDI 712

```

```

Qy 541 ALLELEKPAEYSSMVBPICLPDASHVPAGKAIWWTGWTGHTQYCGTGALILQKGEIRVIN 600
Db 713 ALLELEKSVESYSVVAPICLPDATHVPFPAKAIWWTGWTGHTKGGGAGLILQKGEIRVIN 772
Qy 601 OTTCENILPQOITPRMCMGFLSGVSDSCGDSGCGPLSVYADGRIPQAGVSWGDCQAQ 660
Db 773 OTTCEDILMPQOITPRMCMGFLSGVSDSCGDSGCGPLSVYADGRIPQAGVSWGDCQAQ 832
Qy 661 RNRKGVYTRPLFRDWIKENTGV 683
Db 833 RNRKGVYTRPLVVRDWIKENTGV 855

RESULT 3
BAC35125
ID BAC35125 PRELIMINARY; PRT; 855 AA.
AC BAC35125;
DT 14-APR-2004 (TEMBLrel. 27, Created)
DT 14-APR-2004 (TEMBLrel. 27, Last sequence update)
DT 14-APR-2004 (TEMBLrel. 27, Last annotation update)
DE 0 day neonate kidney cDNA, RIKEN full-length enriched library.
DE clone: D630041B06 product: suppression of tumorigenicity 14 (colon
DE carcinoma), full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=22354663; PubMed=12466851;
RA The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=9279253; PubMed=10349636;
RA Carninci P.; Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=2049374; PubMed=11042159;
RA Carninci P.; Shibata Y.; Hayatsu N.; Sugahara Y.; Shibata K.; Itoh M.;
RA Kono H.; Okazaki Y.; Muramatsu M.; Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=20530913; PubMed=11076661;
RA Shibata K.; Itoh M.; Aizawa K.; Nagaoka S.; Sasaki N.; Carninci P.;
RA Kono H.; Akiyama U.; Nishi K.; Kitsuai T.; Tashiro H.; Itoh M.;
RA Sumi N.; Ishii Y.; Nakamura S.; Hazama M.; Nishine T.; Harada A.;
RA Yamamoto R.; Matsumoto H.; Sakaguchi S.; Ikegami T.; Kashiwagi K.;
RA Fujiwara S.; Inoue K.; Togawa Y.; Izawa Y.; Ohara E.; Watahiki M.;
RA Yoneda Y.; Ishikawa T.; Ozawa K.; Tanaka T.; Matsura S.; Kawai J.;
RA Okazaki Y.; Muramatsu M.; Inoue Y.; Kira A.; Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";

```



RL Genome Res. 10:1757-1771 (2000).  
 RN [6]  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagata T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hori F., Imocani K., Ishii Y., Itoh M., Kagawa I., Kanakawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Saeki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Yamamatsu M., Hayashizaki Y.,  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK052738; BAC35125.1; -  
 SQ SEQUENCE 855 AA; 94654 MW; 4F10E84DA2146DD5 CRC64;

Query Match 84.4%; Score 3209; DB 2; Length 855;  
 Best Local Similarity 83.2%; Pred. No. 2.3e-217;  
 Matches 568; Conservative 59; Mismatches 56; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPARSLKSFVTSVVAFPDSTKVORTDQNSCSFGLHARGVELMRFTTPG 60  
 DB 173 MAVERVVTLPARRALKSFVLTSVVAFPIDPRMLQRTDQNSCFALHARGAAYTRFTTPG 232  
 QY 61 PPDSPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERSGLVTVYNTLSPEMHPALVQ 120  
 DB 233 PPNSPYPAHARCOMVLRGDADSVLSLTFRSFDVAPCDHSGSDLVTVYDLSLSPMEHPAVVR 292  
 QY 121 LCGTPPSYNTLTFHSSQNVLLITLITNTERRHGFEATFQLPKMSCGGLRLKAQGTEN 180  
 DB 293 LCGTSPSYNTLTFSSQNVFLVTLITNDRHPGFATFQLPKMSCGGLSLDTQGFPS 352  
 QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLLEPRRACGTCPKDYVINGEKYCGE 240  
 DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLLEPRRACGTCPKDYVINGEKYCGE 412  
 QY 241 RSQFVTVSNKNTVRPHSDSYDTDTGFLAELYSYSSDPGPGFTCRGTGRCIKELRCD 300  
 DB 413 RSQFVTVSNKNTVRPHSDSYDTDTGFLAELYSYSSDPGPGFMCKTGRCIKELRCD 472  
 QY 301 GMADCTHSDLELNCSCAGHOFCTCKNRCRKLFWVCDSVNDCCGNSDEGSCSPAGQFRC 360  
 DB 473 GMADCPYSDERYCRCAATHQFTCKNQCKPLFWVCDSVNDCCGNSDEGSCSPAGSFKC 532  
 QY 361 SNGKCLSKSQCGKNDGDSDEASCPKVVVYCTKHTYRCLNGLCLSKGNPECDGKED 420  
 DB 533 SNGKCLFQSQCKGKNDGDSDBASCDVVVVGCTKHTYRCLNGLCLSKGNPECDGKED 592  
 QY 421 CSDSDSEKDCDCCGLRSFTQARVVGTDADGEMPMQVSLHALQGHITCGASLISPMWLV 480  
 DB 593 CSDSDSEKDCDCCGLRSFTQARVVGTDADGEMPMQVSLHALQGHITCGASLISPMWLV 652  
 QY 481 SAACGYIDDRFRSDPTQWTAFLGHDGORSAPRGVERLKLITSHFPNDFTFYDI 540  
 DB 653 SAACGFODDKFKKYSDTYMTAFLGLDQSRASAGVELLKRITHTPSFNDFTFYDI 712  
 QY 541 ALLELEKPAEYSSAWRPICLPDASHVFPAGRAIWTGNGHTQYGGTGLALIQKEIRIYN 600  
 DB 713 ALLELEKSVESYTVRPLCPDATHVFPAGRAIWTGNGHTKRGSTGLALIQKEIRIYN 772  
 QY 601 QTTCENTLLPQGITPRMNCVGLSGGVDSCGSDSGGSLSSVADRIIFQAGVSWGDCGQ 660  
 DB 773 QTTCEDLMPQGITPRMNCVGLSGGVDSCGSDSGGSLSSVADRIIFQAGVSWGDCGQ 832  
 QY 661 RNKPGVYTRRLPLFRDWIKENTGV 683  
 DB 833 RNKPGVYTRRLPLFRDWIKENTGV 855

RESULT 4

09JUL7  
 ID 09JUL7 PRELIMINARY; PRT; 855 AA.  
 AC 09JUL7  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Membrane bound serine protease (Membrane bound arginine specific  
 DE serine protease).  
 GN Name=MSP,  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Jejunum;  
 RA MEDLINE=21458307; PubMed=11573963;  
 RA Sacomi S., Yamazaki Y., Tsuzuki S., Hitomi Y., Iwanaga T., Fushiki T.,  
 RT "A role for membrane-type serine protease (MT-Sp) in intestinal  
 RT epithelial turnover."  
 RL Biochem. Biophys. Res. Commun. 287:995-1002 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Duoenum;  
 RA Inoue H., Takahashi K., Kishi K.,  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1. SIMILARITY: Belongs to peptidase family 11.  
 DR EMBL; AB037898; BAB03502.1; -  
 DR PIR; JC7731; JC7731.  
 DR HSSP; P00760; 1EZX.  
 DR MEROPS; S01.302; -  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR002172; LDL\_receptor\_A.  
 DR InterPro; IPR001254; peptidase\_S1.  
 DR InterPro; IPR001314; peptidase\_S1A.  
 DR InterPro; IPR009003; Pept\_Ser\_Cys.  
 DR Pfam; PF00431; CUB; 2.  
 DR Pfam; PF00057; Ldl\_recept\_a; 4.  
 DR Pfam; PF00089; Trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR PROSITE; PS01180; CUB; 2.  
 DR PROSITE; PS01209; LDLRA\_1; 2.  
 DR PROSITE; PS00068; LDLRA\_2; 4.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 855 AA; 94955 MW; 35806B7ECFCF03D CRC64;

Query Match 84.4%; Score 3197; DB 2; Length 855;  
 Best Local Similarity 83.0%; Pred. No. 1.6e-216;  
 Matches 567; Conservative 59; Mismatches 57; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPARSLKSFVTSVVAFPDSTKVORTDQNSCSFGLHARGVELMRFTTPG 60  
 DB 173 MAVERVVTLPARRALKSFVLTSVVAFPIDPRMLQRTDQNSCFALHARGAAYTRFTTPG 232  
 QY 61 PPDSPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERSGLVTVYNTLSPEMHPALVQ 120  
 DB 233 PPNSPYPAHARCOMVLRGDADSVLSLTFRSFDVAPCDHSGSDLVTVYDLSLSPMEHPAVVR 292  
 QY 121 LCGTPPSYNTLTFHSSQNVLLITLITNTERRHGFEATFQLPKMSCGGLRLKAQGTEN 180  
 DB 293 LCGTSPSYNTLTFSSQNVFLVTLITNDRHPGFATFQLPKMSCGGLSLDTQGFPS 352  
 QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLLEPRRACGTCPKDYVINGEKYCGE 240

```
Dh 353 SPYYPGHPNINCTWNIKVPNNNNKVRKFLFYLVDPNIPVSGCTKDYVEINKEKCGE 412
Qy 241 RSQFVTSNSNKTIVRFRHSQSYTDTGFLAEYLSYSDSDPCPGQFTCRGTGRCKRELKCD 300
Dh 413 RSQFVTSNSNKTIVRFRHSQSYTDTGFLAEYLSYSDSDPCPGQFTCRGTGRCKRELKCD 472
Qy 301 GMADCTHSDBLNCSGAGHOFTCKNFKCKPLFVVCDSVNDGDSNDQSCSPAQTFRC 360
Dh 473 GMADCPYSDERHRCNATQFMCKNQFCCKPLFVVCDSVNDGDSNDQSCSPAQTFRC 532
Qy 361 SNGKLSKSOQCNKDCGSDGSDASCPKVVVCTHTYRCNLGLSKNPECDKED 420
Dh 533 SNGKLSKSOQCNKDCGSDGSDASCDVNAVSCITYRCQGLCKNPECDKED 592
Qy 421 CSQSDSEKDCDCGSRFTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480
Dh 593 CSQSDSEKDCDCGSRFTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652
Qy 481 SAACHYIDRGRFRYSDPTQWTAFLGLHDGQSRAPGVQERLKRILSHPFNDFPDYDI 540
Dh 653 SAACHFQDETFKYSDDHTMTAFLGLHDGQSRASGVQENHLKRILSHPFNDFPDYDI 712
Qy 541 ALLELEKPAEYSSWVRPCLPDASHVPRAGKATVMTGHTQVGTGALLQKELIVIN 600
Dh 713 ALLELEKPAEYSTVVRPCLPDNTHVPAGKALVMTGHTGEGTALLQKELIVIN 772
Qy 601 QTTCEMLLPQOITPRMMCVFGLSGGVSCGSDSGPLSYEADGRIFQAGVWSGDDCAQ 660
Dh 773 QTTCEMLLPQOITPRMMCVFGLSGGVSCGSDSGPLSYEADGRIFQAGVWSGDDCAQ 832
Qy 661 RNKPGVYTRILPLFDMIKENTGV 683
Dh 833 RNKPGVYTRILPEVDMIKEQTV 855

RESULT 5
ID 08WVC1 PRELIMINARY; PRT; 422 AA.
AC 08WVC1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE S114 protein (Fragment).
GN S114;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefter C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hale F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Tohiyuki S., Carninci P., Prange C.J.,
RA Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
```

```
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBS databases.
CC -1- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; BC018146; ARI18146.1; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001314; Peptidase_S1.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00089; Trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 4.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS01209; LDLA_1; 2.
DR PROSITE; PS50068; LDLA_2; 4.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR Hydrolase; Hydrolase; Serine protease.
FT NON_TER
SQ SEQUENCE 422 AA; 46257 MW; 2C99875D1B58B319 CRC64;

Query Match 62.8%; Score 2379; DB 2; Length 422;
Best Local Similarity 100.0%; Pred. No.2.8e-159;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 262 SYTDGFLAEYLSYSDSDPCPGQFTCRGTGRCKELRCGMDCTHSDBLNCSGAGHQ 321
Dh 1 SYTDGFLAEYLSYSDSDPCPGQFTCRGTGRCKELRCGMDCTHSDBLNCSGAGHQ 60
Qy 322 FTCKNFKCKPLFVVCDSVNDGDSNDQSCSPAQTFRCNGKLSKSOQCNKDCGSG 381
Dh 61 FTCKNFKCKPLFVVCDSVNDGDSNDQSCSPAQTFRCNGKLSKSOQCNKDCGSG 120
Qy 382 SDSAQCPKVVVCTHTYRCNLGLCLSKNPECDKEDCSQSDSDKDCGSRFTROA 441
Dh 121 SDSAQCPKVVVCTHTYRCNLGLCLSKNPECDKEDCSQSDSDKDCGSRFTROA 180
Qy 442 RVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLVSAACHYIDRGRFRYSDPTQWT 501
Dh 181 RVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLVSAACHYIDRGRFRYSDPTQWT 240
Qy 502 AFLGLHDGQSRAPGVQERLKRILSHPFNDFPDYDIALLEKPAEYSSWVRPCLP 561
Dh 241 AFLGLHDGQSRAPGVQERLKRILSHPFNDFPDYDIALLEKPAEYSSWVRPCLP 300
Qy 562 DASHVPRAGKALVMTGHTQVGTGALLQKELIVINQTTCEMLLPQOITPRMMCVGF 621
Dh 301 DASHVPRAGKALVMTGHTQVGTGALLQKELIVINQTTCEMLLPQOITPRMMCVGF 360
Qy 622 LSGGVSCGSDSGPLSYEADGRIFQAGVWSGDDCAQRNKGVTYTRILPLFDMIKENT 681
Dh 361 LSGGVSCGSDSGPLSYEADGRIFQAGVWSGDDCAQRNKGVTYTRILPLFDMIKENT 420
Qy 682 GV 683
Dh 421 GV 422

RESULT 6
ID 06GR54 PRELIMINARY; PRT; 845 AA.
AC 06GR54;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE S114-A-prov protein.
```

GN Name=sl4-A-prov;  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 OC Xenopodidae; Xenopus.  
 CX NCBI\_TaxId=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Kaiser R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Alechul S.F., Zeeberg B., Buetow K.H., Scheer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stepieton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Umed T.B., Toshiyuki S., Carrinci P., Prange C.,  
 RA Rata S.S., Loggiano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,  
 RA Bock S.A., McGwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
 RA Krzywinski M.I., Skalek U., Smallos D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Maitra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22341132; PubMed=12454917;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.,  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT Initiative".  
 RL Dev. Dyn. 225:384-391(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.,  
 RL Submitted (May-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to peptidase family S1.  
 DR EMBL; BC071077; AAH71077.1; -.  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR002172; LDL\_receptor\_A.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR InterPro; IPR009003; Pept\_Ser\_Cys.  
 DR Pfam; PF00431; CUB; 2.  
 DR Pfam; PF00057; Ldl\_recept\_a; 4.  
 DR Pfam; PF00089; Trypsin; 1.  
 DR PRINTS; PRO0722; CHYMOTRYPSIN.  
 DR PRINTS; PRO0261; LDLRECEPTOR.  
 DR SMART; SM00042; CUB; 2.  
 DR SMART; SM00192; LDLa; 4.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS01180; CUB; 2.  
 DR PROSITE; PS01209; LDLRA\_1; 2.  
 DR PROSITE; PS00068; LDLRA\_2; 4.  
 DR PROSITE; PS02440; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM HydroLase; Protease; Serine protease.  
 SQ SEQUENCE 845 AA; 93593 MW; B917386C95BC73BD CRC64;

Query Match 60.2%; Score 2280.5; DB 2; Length 845;  
 Beel Local Similarity 57.3%; Pred. No. 5.3e-152;  
 Matches 393; Conservative 125; Mismatches 161; Indels 7; Gaps 4;

1 MAERVVMLPFRASLKSFFVTSTVAFPTDSKTVQRTDNCSPGLHARGVLEWRFTTPG 60

Db 164 ISELKLPSTVNROR---TFALDSLVAVPTDPOQLARVFNKSSCAVFLHSSNGVAKFFSPG 220  
 Qy 61 PPSPSPYAHARCOMALRGDADSVLSLTFRSPDLASCDERSDLVYNTLSPMEPHALVQ 120  
 Db 221 PPSPSPYAHARCOMALRGDADSVLSLTFRSPDLASCDERSDLVYNTLSPMEPHALVQ 280  
 Qy 121 LCGTYPSPYAHARCOMALRGDADSVLSLTFRSPDLASCDERSDLVYNTLSPMEPHALVQ 180  
 Db 281 LCGTYPSPYAHARCOMALRGDADSVLSLTFRSPDLASCDERSDLVYNTLSPMEPHALVQ 340  
 Qy 181 SPYYPGHYPNNIDCTWNIIEVNNQHVKVRPFYLLLEPRRACGTCPPDYVEINGEKYCGE 240  
 Db 341 SPYYPGHYPNNIDCTWNIIEVNNQHVKVRPFYLLLEPRRACGTCPPDYVEINGEKYCGE 400  
 Qy 241 RSQFVTSNKNKTTVRHSQSTTDTGFLAEVLSYSDSDCPQGFTRTRCRKEARCD 300  
 Db 401 KEFVVSNNSSKMSVVRVSDSYTDGFTAEVLSYEBRNCPPOFTCRSGRCIRLDQKCD 460  
 Qy 301 GMADCTHSDMLNCSODAGHOFCKN-KFCKPLFWVCDVNDGSDNDEGSCSPAOFTFR 359  
 Db 461 GMADCTHSDMLNCSODAGHOFCKN-KFCKPLFWVCDVNDGSDNDEGSCSPAOFTFR 519  
 Qy 360 CSNGKCLSKSQCGKDDCGSDGDBASCPRVNVYCTKTYRCLNGCLSKGNPCDGKE 419  
 Db 520 CGNGKCI PDSQKCDRVNCGDGDDEACDQVLTACTEYVYKCKNNQCTKKNPCDGEN 579  
 Qy 420 DCSDSGSEK--DCDCGLRSTRQARVVGCTDABEGEMPMOVSILHALGQGHICASLSRN 477  
 Db 580 DCSDSGSEK--DCDCGLRSTRQARVVGCTDABEGEMPMOVSILHALGQGHICASLSRN 639  
 Qy 478 MLVSAACHYIDDRGFRSDPTQMTAFGLHDOSORSAPOVERBLKRIISHPFNDPTFD 537  
 Db 640 MLVSAACHYIDDRGFRSDPTQMTAFGLHDOSORSAPOVERBLKRIISHPFNDPTFD 699  
 Qy 538 YDIALLELEKPAEYSSWVRPCLPDASHVPPAKAIWVTMGHTQYGTGALILQKGEIR 597  
 Db 700 YDIALLELEKPAEYSSWVRPCLPDASHVPPAKAIWVTMGHTQYGTGALILQKGEIR 759  
 Qy 598 VINOTTEENILPQOITPRMNCVGFSLSGVDSCGSDSGPLSYBADGRITQAGVSVSGDG 657  
 Db 760 VINOTTEENILPQOITPRMNCVGFSLSGVDSCGSDSGPLSYBADGRITQAGVSVSGDG 819  
 Qy 658 CAORNKPGVYTRLPFLPDWIKENTGV 683  
 Db 820 CAORNKPGVYTRLPFLPDWIKENTGV 845

RESULT 7  
 Q9DGR1 PRELIMINARY; PRT; 845 AA.  
 AC Q9DGR1;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Homology of human MT-Spl.  
 GN Name=XMT-Spl.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 OC Xenopodidae; Xenopus.  
 CX NCBI\_TaxId=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20363741; PubMed=10903452;  
 RA Yamada K., Takabatake T., Takeshima K.,  
 RT "Isolation and characterization of three novel serine protease genes  
 RT from Xenopus laevis".  
 RL Gene 252:209-216(2000).  
 CC -1- SIMILARITY: Belongs to peptidase family S1.  
 DR EMBL; AB038498; BAB08218.1; -.  
 DR HSSP; P00760; 38TH.  
 DR MEROPS; S01.050; -.

DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
DR GO; GO:0004295; F:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR000859; CUB.  
DR InterPro: IPR002172; LDL\_receptor\_A.  
DR InterPro: IPR001254; Peptidase\_SI.  
DR InterPro: IPR001314; Peptidase\_SIA.  
DR InterPro: IPR009003; Pept\_Ser\_Cys.  
DR Pfam: PF00431; CUB; 2.  
DR Pfam: PF00057; Ldl\_recept\_a; 4.  
DR Pfam: PF00089; Trypsin\_1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR PRINTS: PR00261; LDLRECEPTOR.  
DR SMART: SM00042; CUB; 2.  
DR SMART: SM00192; IDLA; 4.  
DR SMART: SM00020; TRYP\_SPE; 1.  
DR PROSITE: PS01180; CUB; 2.  
DR PROSITE: PS01209; LDLRA\_1; 2.  
DR PROSITE: PS50068; LDLRA\_2; 4.  
DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
DR PROSITE: PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
DR PROSITE: PS00135; TRYPsin\_SER; 1.  
KW Hydrolyase; Protease; Serine protease.  
SQ SEQUENCE 845 AA; 93597 MW; 7FD7E2651A758B7 CRC64;

Query Match 59.5%; Score 2254.5; DB 2; Length 845;  
Best Local Similarity 57.0%; Pred. No. 3.6e-150;  
Matches 391; Conservative 123; Mismatches 165; Indels 7; Gaps 4;

QY 1 MAERVVMLPPRRASLSFVVTSVAPPTDSKYQKRDNSCSGGLHARGVELMRFTTPG 60  
DB 164 ISELKLPSPVNROR---TFALDSILVAIPDQIARVFNKSCAYFLHSSNGVAKFSSPG 220  
QY 61 PPDSPPVARRCOMALRGDSDVSLTFPSFDLASCERSGDLVTYVNTLSPEMPLATQ 120  
DB 221 PPDSPPVARRCOMALRGDSDVSLTFPSFDLASCERSGDLVTYVNTLSPEMPLATQ 120  
QY 121 LCGTPPSYNLTFHSONVLLITLITERRHPGFATFQLPMSGGRRLKRAQSTFN 180  
DB 281 LCGIYPPSYNLTFHSSNMVLTIVTDNVGKFPGLAFBQKLPRTSLCGGLIRDSGFIT 340  
QY 181 SPYRGHYPRINDCTNMIIEVNNHNVKRFYFLLEPRACGCPDYVEINERKCGE 240  
DB 341 SPYFPAPHYPTSTESIMDIQVDPNKFVRFNMFYLABPGVPVTKCTDPVEIKQKCGE 400  
QY 241 RSQFVTVSNNSKTIIVRFHSDQSYDTGTFLAEYLSYDSDPCPGQFTCRGRCIKELRCD 300  
DB 401 KEFVTVSNNSKMSKVRFSVDSQSYDTGTFLAEYLSYBERNFCPDQFTCRGRCIKELRCD 460  
QY 301 GWADCTDHSDELNCSGAGHQTCKN-KFCKPLFWCDSVNDGSDNSDEQSCGPAQTFR 359  
DB 461 GWNDCEBFDSEMSCTCTA-LQFRCVNSKLCRPSYFICDGVNDGSDSDSLACCKPNNTPK 519  
QY 360 CSNCKCLSKSQGCKGDCDGDGSDASCPKYNVNTCTKHRYRCNLGCLSKNGNECCKGE 419  
DB 519 CGNCKCLPSQCKCRVNCGDSDEACDQVLTACTTCEYTYCKCNKNCCTIKKPECCGEN 579  
QY 520 DCSGDSDEK--DCDGLRSFTQARVVGTDADGEMPMQVSLHALQCGHICGASLSPN 477  
DB 477 DCSGDSDEK--DCDGLRSFTQARVVGTDADGEMPMQVSLHALQCGHICGASLSPN 477  
QY 580 DCSGDSDEK--DCDGLRSFTQARVVGTDADGEMPMQVSLHALQCGHICGASLSPN 477  
DB 477 DCSGDSDEK--DCDGLRSFTQARVVGTDADGEMPMQVSLHALQCGHICGASLSPN 477  
QY 538 YDIALLELEKPAEISSWVRPCLPDASHVPAPGAAIIVTGWGHTQYQVGTGALLIQKEIR 597  
DB 597 YDIALLELEKPAEISSWVRPCLPDASHVPAPGAAIIVTGWGHTQYQVGTGALLIQKEIR 597  
QY 700 NDIAVLELEKPEVETDRIQVCPIDSETHDPVGPPIWVGALKEGGAAVILIQKAEIR 759  
DB 759 NDIAVLELEKPEVETDRIQVCPIDSETHDPVGPPIWVGALKEGGAAVILIQKAEIR 759  
QY 598 VINQTEENLLPQITPRMCKVGLSGVDSGQSDSGPLSSVADARIRFOAGVAVSGDG 657  
DB 657 VINQTEENLLPQITPRMCKVGLSGVDSGQSDSGPLSSVADARIRFOAGVAVSGDG 657  
QY 760 IINQTEENLLPQITPRMCKVGLSGVDSGQSDSGPLSSVADARIRFOAGVAVSGDG 819  
DB 819 IINQTEENLLPQITPRMCKVGLSGVDSGQSDSGPLSSVADARIRFOAGVAVSGDG 819

QY 658 CAQRNKGYYTRLPFRDWIKENTGV 683  
DB 820 CARRNKPGRYTVKVMRDSKDTGL 845

## RESULT 8

Q6DEV0 PRELIMINARY; PRT; 663 AA.

AC Q6DEV0; 01-OCT-2004 (TrEMBLrel. 28, Created)  
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Hypothetical protein.  
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodidae; Xenopus.  
OX NCBI\_TaxID=8364;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=2238257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusik A.K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshitoki S., Carninci P., Prange C.,  
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Miliady S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Green E.D., Dickson M.C.,  
RA Blakeley R.W., Touchman J.W., Shevchenko Y., Bouffard G.G.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzyzanski M.I., Skalska U., Smalins D.E., Schmeich A., Schell J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RA Klein S., Strausberg R.;  
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC076994; AAH76994.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 663 AA; 73914 MW; 46B2A56C657C7739 CRC64;

Query Match 40.7%; Score 1543; DB 2; Length 663;  
Best Local Similarity 41.5%; Pred. No. 3.3e-100;  
Matches 278; Conservative 91; Mismatches 115; Indels 186; Gaps 4;

QY 17 KSFVTVSVVAFPTDSVTQRTQDNCSPGLHARGVELMRFTTPGPPSPVARRCOMAL 76  
DB 177 RAFAVSLVAYPPDPIANFNKSSCAFLHSSAGVMTKSSSPGPDTPVPPVARRCOMAL 236  
QY 77 RGADAVSLTFPSFPLASCDBERGSDLVYVNTLSPEMPLATQVQLGTPPSYNLTFHSS 136  
DB 237 RADAGMIRLKEFTPFMEKCKANAGDFVWYDLSLIEPRAQRLCGIYPPSYNLTFHSS 296  
QY 137 QNVLTLLITNTRRRHGFATFQLPMSGGRRLKRAQSTFN 180  
DB 297 SNVMTLVITDVGKFPGLAEYLSYDSDPCPGQFTCRGRCIKELRCD 460  
QY 197 NIEVPNNQHYKVRKFPYLLIEPRACGCPKDYVEINERKCGEISQFVTVSNNSKTIIVR 256  
DB 357 DIQVDPNKFVRLAFNMFYLABPGVPVTKCTDPVEIKQKCGE 400  
QY 257 FHDSDSYDTGTFLAEYLSYDSDPCPGQFTCRGRCIKELRCDGADCTDHSDELNCSG 316  
DB 316 FHDSDSYDTGTFLAEYLSYDSDPCPGQFTCRGRCIKELRCDGADCTDHSDELNCSG 316

```
Db 417 FVSDQSYTDGTGTAETLYSEYPRNCPDQFACKSGRCIRLIDKCDGMNDCEDPSDEKSTC 476
Qy 317 DAGHQFCKN-KFKCPPLFWCDVNDGSDNSDEGSCSPACTPFCNSGKCLSKSQCCNGK 375
Db 477 TA-LQFCTNLSKLPSTFVCDGVNDGSDSDLAQCQPNNTYKCGNGKCLPSQKCDRT 535
Qy 376 DDCDGDSDDEASCPRNVVNTCTKHTYRCINGLCLSKGNPECDGKEDSDSDSDE--KDCDCG 433
Db 536 DNGDGDSDDEASCGRVLTLTTCETRYTKCNNGCITRKNPCEGENDCGSDSEISAKNCG 595
Qy 434 LRSTQARVVGTDADGEMWPMOVSILHALQCGHI CGASLISPMWLYSAHCHYIDRGR 493
Db 596 KRPFCKSRIVGQVNA----- 611
Qy 494 YSDPTQMTAFGLGHDGSRASPGVERLKRILSHPFENDPTFYDIALLEKPAEYSS 553
Db 612 ----- 611
Qy 554 MVRPCLPDASHVPAGKAIWVTGMHTQYGTGALILQKGEIRVINGTCENLLPOQIT 613
Db 612 ----- 611
Qy 614 PRMVCVPLSGVDSCGDSGGLSVADGRIFQAGVNSGDCQAQRNFGVTRLPLE 673
Db 612 -----DSGGLSVSELNKNKYLAVIGVSGEGCARRNKGVTYTRVAM 653
Qy 674 RDMKENTGV 683
Db 654 RDMTRDTGL 663

RESULT 9
QRTY8 PRELIMINARY; PRT; 572 AA.
AC QRTY8;
DT 01-MAR-2004 (TREMblrel. 26, Created)
DT 01-MAR-2004 (TREMblrel. 26, Last sequence update)
DT 01-OCT-2004 (TREMblrel. 28, Last annotation update)
DE Type II transmembrane serine protease 7 precursor (Hypotheetical
protein FLJ16088).
GN Name=TPRSS7;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22722134; PubMed=12838346;
RA Puente X.S., Sanchez L.M., Overall C.M., Lopez-Otin C.;
RT "Human and mouse proteases: a comparative genomic approach.";
RL Nat. Rev. Genet. 4:544-558(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Nimomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuya N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Matsui Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Katsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Katsuhori K., Takahashi-Fuji A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masubo Y., Nagai K., Isogai T.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
EMBL: BN000125; CAD67577.1; -.
DR MEROPS; S01.072; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004235; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
```

```
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00057; LDL_recept_a; 2.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease; Signal; Transmembrane.
FT SIGNAL 1 23
FT CHAIN 24 572 type II transmembrane serine protease 7.
SQ SEQUENCE 572 AA; 64041 MW; 0458CBEA996BA285 CRC64;

Query Match 26.9%; Score 1017.5; DB 2; Length 572;
Best Local Similarity 36.3%; Pred. No. 2.8e-63;
Matches 226; Conservative 98; Mismatches 223; Indels 75; Gaps 21;

Qy 72 CQNALRDADSVSLTRFSFDLASCDERGSGLVTVNTLSPEMHALVOLCTTPPSYNL 131
Db 2 CPELVAVIGVLTIRLSIKSIQI-EADNCVDSLTLYDSLPIRSSILYRIC-EPTRTLM 58
Qy 132 TFSQNVLLITLTITNERBHPGEATFPQLPMHSSCGRLRK---AQTSPSPYRGH 187
Db 59 SFVSTNNMLVTFKSPHRIKSLGRATFVYIPECKCENTVAVKDITFEGKISSPYPSY 118
Qy 188 YPNPIDCTNANIEVNNQHVVRKFFYLPRRACGTCPDYVEINKEKYCGE--RSQFV 245
Db 119 YPKCKCTWKFQRT-SLTSLGALKFNVYSITKSKMCKEHWMEINHWCGSYMHDQTI 177
Qy 246 VTSNSNKITVRFSHSDSYDTGTFLAETLYSDSDPCF-GQFTCTGTCIKELRCQGMAD 304
Db 178 FRVPSPLVHIQLQCSSRLSDKPLAEVGSYNIISQPCVGSFRSSGGLCVPAQRCQDVND 237
Qy 305 CTHSDSELNCSQAGHQFTCKNPKC---KPLFWVCDSVNDGSDSDGSCSPAQTRFCS 361
Db 238 CFDSDEL-----FCVSPDP-----ACNTSFR-Q 261
Qy 362 NGKCLASQOCCNGKDDGSDSDDEASCPRNVVNTCTKHTYRCINGLCLSKGNPECDGKEDC 421
Db 262 HGPLI-----CDGRDDENRDENGNC--TQSIQNNRTFPGMNDICRQKNAKCDGVDC 314
Qy 422 SDGSDKDCDCGLRSFTRQARVVGTDADGEMWPMOVSILHALQCGHI CGASLISPMWLV 481
Db 315 PDGSDDEGCICS-RSSSALHRIIGTDTLEGWPMQVSLHFVGSAY-CGASVISREWLIS 372
Qy 482 AAHCYIDRGRFSRSDPTQMTAFGLGHDGSRASPGVERLKRILSHPFENDPTFYD 539
Db 373 AAHCF--HGNRLSDPTQMTAFGLGHDGSRASPGVERLKRILSHPFENDPTFYD 423
Qy 540 IALTELE--PRAEYSWVRPCLPDASHVPAGKAIWVTGMHT-QYGTGALILQKGEI 596
Db 424 IALLQLSIANPRTKQLQIGICIPRQGRVSEKCKVITMGKRNHEMDNNGSLVLQDAEV 483
Qy 597 RVINGTCENLLPQITPRMVCVPLSGVDSCGDSGGLSVADGRIFQAGVNSGDCQAQRN 655
Db 484 ELIDQTLQVSTY-GIITSRLMAGINSGRDACKGSGGGLSGRKSQDKWIIILGIYSWG 542
Qy 656 DGCAGRNKPGVYTRLPFRDWI 677
Db 543 HGCGRNPGVYTRVSNFPMI 564

RESULT 10
BAD18401 PRELIMINARY; PRT; 572 AA.
AC BAD18401;
DT 12-MAY-2004 (TREMblrel. 27, Created)
DT 12-MAY-2004 (TREMblrel. 27, Last sequence update)
```

DT 12-MAY-2004 (Tremblrel. 27, last annotation update)  
 DE CNNA FLJ16088 f1s, clone NT2RP7008435, weakly similar to EPITTHIN (EC 3.4.21.-).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 ON NCBI\_taxonomy=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Niinomiya K., Magatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,  
 RA Kusuda N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie K.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isegai T.,  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AKJ1311; BAD18401.1; -  
 SQ SEQUENCE 572 AA; 64041 MW; 0458CEA996EA285 CRC64;

Query Match 26.9%; Score 1017.5; DB 2; Length 572;  
 Best Local Similarity 36.3%; Pred. No. 2.8e-63;  
 Matches 226; Conservative 98; Mismatches 223; Indels 75; Gaps 21;

QY 72 CQWALRGDADSVLSLTPRFDLASCDERGSDLVTVTLSLSPMEPHALVOLGCTYPSPYNL 131  
 DB 2 CHTPLVAIVGLVRLSLKSIQI-BADNCVDSLTITVSLPIRSILYRIC-EPTITLM 58  
 QY 132 TFFSSQVLLITLTTERHPGEATFQPLPMSSCGRLRK--AQGTSPYPPGH 187  
 DB 59 SFVSTNNMLVTFKSPHRLSGIRAFEVYPEQKCNVTLVKQITFEGLTSSPYPSY 118  
 QY 188 YPPNIDCTNINVENNHNQVAVKFEFLBPRACGCPDYVINEKXGE--RSQFV 245  
 DB 119 YPPCKCTKMPQT-SLSTLGLALFVNYSTKSKCKCHGMMWLNHMTGSGYMDHOTI 177  
 QY 246 VTSNSNKTIVRFHSDQSYDTGFLAEVLSYDSNPPC-GQFTCTGRCIKELACDGMAD 304  
 DB 178 FRVPSPLVHQLQSSSLSDKPLAEGSYNISQPCVSPFRSSGLCVQAQRCDVDND 237  
 QY 305 CTHSDELNCSADGHOFTCKNFKC--KPLFWVCDSVNDGNSDEQSCSPAQTRFCS 361  
 DB 238 CPDESDEL-----FCVSPPOP-----ACNTSSPR-Q 261  
 QY 362 NGKLSLSSQCGNGKDDGSDSCSKRVAVVCTKHTVCLNGCLSKNGPEDGKEDC 421  
 DB 262 HGPLT-----CDGRDDENGRDEBQNC--TOSIPCNNTFTKCGNDICFRKONAKCDGTVD 314  
 QY 422 SDGSDKDCDCGLRSFTROARVVGTDADRGEMFQVSLHALCGGHIAGSLISPMVLVS 481  
 DB 315 PDGSDGEGCTCS-RSSSALHRIITGDTTLBGGMWQVSLHFGVSAV-CGASVISREMLLS 372  
 QY 482 AAHCYIDRGRRSDPLQWTAFLGLHDQSQRS--APGVQERRLKRIISHPENDFTFYD 539  
 DB 373 AAHCF--HGRRLSDPTPMTAHLGMYVQGNAKFVSP-----VRIIVHEVYNSQTFDYD 423  
 QY 540 IALLLEF--KPAEYSSWVRPCLPDASHVPRAGALIVWTGHT-QVCGGALILQGEI 596  
 DB 424 IALLQLSIAWPEITKOLIPICIPPTGORVRSYGKCVWTGRRHEDNKGSLVLAQAEV 483  
 QY 597 RVINQTTCCENLIPQITPRMVCVFLSGVDS CGGSDGSPLS--SVEADGRIFQAGVSWG 655  
 DB 484 ELIDQITCVSTY-GIITSRLCAGIMSGKXCKDCKGSDGSPLSCKRSDKMLILGIYSWG 542  
 QY 656 DGCQRNKPQVYTRLPFRDWT 677  
 DB 543 HGGGRPNFPQVYTRVSNFVPMI 564

RESULT 11  
 08BIT6

ID 08BIT6 PRELIMINARY; PRT; 572 AA.  
 AC 08BIT6;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, last sequence update)  
 DT 01-MAR-2004 (Tremblrel. 26, last annotation update)  
 DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length  
 DE enriched library, clone:B230219123 product:weakly similar to BLOOD  
 DE COAGULATION FACTOR XI.  
 GN Name=B230219123Rik;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_taxonomy=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RC MEDLINE=21085660; PubMed=11217851;  
 RX RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RC THE FANTOM Consortium;  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RC MEDLINE=20499374; PubMed=11042159;  
 RX Kancinci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kanno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes."  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RC MEDLINE=20530913; PubMed=11076861;  
 RX Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kanno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
 RA Yamamoto R., Matsunaga H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujisaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Matsubara M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RC Aichi U., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozawa T.,  
 RA Hori F., Imclani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katon H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
 RA Kurinara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Osakato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takada Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

-1- SIMILARITY: Belongs to peptidase family S1.  
 DR EMBL: AK045663; BAC2448.1; --  
 DR HSSP: P00760; IEZX.  
 DR MEROPS: S01.072; --  
 DR MGD; MG1:2686594; B230219123R1K.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR000859; CUB.  
 DR InterPro: IPR002172; LDL\_receptor\_A.  
 DR InterPro: IPR001254; Peptidase\_S1.  
 DR InterPro: IPR001314; Peptidase\_S1A.  
 DR InterPro: IPR009003; Pept\_Ser\_Cys.  
 DR Pfam; PF00431; CUB; 1.  
 DR Pfam; PF00057; LDL\_recept\_a; 3.  
 DR Pfam; PF00089; Trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR SMART; SM00042; LDLa; 1.  
 DR SMART; SM00192; LDLa; 3.  
 DR SMART; SM00020; TRYP\_SPE; 1.  
 DR PROSITE; PS01180; CUB; 2.  
 DR PROSITE; PS01209; LDLRA\_1; 1.  
 DR PROSITE; PS50068; LDLRA\_2; 2.  
 DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 DR Hydrolase; Protease; Serine protease.  
 KW SEQUENCE 572 AA; 63757 MW; 21C0EC52B5F4301 CRC64;  
 SQ  
 Query March 26.8%; Score 1015.5; DB 2; Length 572;  
 Best Local Similarity 36.3%; Pred. No. 3.9e-63;  
 Matches 225; Conservative 98; Mismatches 226; Indels 71; Gaps 21;  
 QY 72 CQNALRGDADSVLSLTFRSPDLASCDERGSIDLTVYNTLSPEMHAVOLCGTYPSPSYNL 131  
 DB 2 CHPFLVAIVGLIRLSTIESIQTL-EADNCITDLSLTVYDLSLIRSAIIVRIC--EPTRTIM 58  
 QY 132 TFHSSQNVLLITLTNTERRRHPGEATFPQLPRNMSCGG-----RLRAQGTNSPYPG 186  
 DB 59 SFVSTNNMLVILKSPYRRLAGIRAYFEVIPR-OKCESTLIVKINSFEKIKSPYPS 117  
 QY 187 HYPNIDICTNIEVYNNQHVAVRPFKFLLEPRACGCPDYVINEKCGE--RSQF 244  
 DB 118 YVPRCKCKTWTFFOT-SLSTLIALKFYNYSTTKSAKCEBGMWEINEMHMGCSYMDHET 176  
 QY 245 VVTSNSNKITVRFHSDSYDTDTGLAEYLSTYSSDPCP-GQFTGRTGRCIRKEILCDGWA 303  
 DB 177 IFRPSPPLVHQLQCSSRLSKPLIVEYGVNISOQCPAGSPRSSGLCTVPAQRCQGVN 236  
 QY 304 DCTDHSDELINCSAAGHOFTCKNFKCKPLFMVVCDSVNDGNSDBQCGSCPAQTFRCGNG 363  
 DB 237 DCFPESDELFCVT-----VKP-----ACNSSSFR-OHG 263  
 QY 364 KCLSKSQQCNCKDQCGSDASGCPKXNVVVTCTKHTVRCNLGLSKENPECDKEDCDSD 423  
 DB 264 PLV-----CDGFRCDEBDQDQNCNR--SLCTSRTEFCGNDICRKNACCDGIVDPPD 316  
 QY 424 GSDKEDCDGSLRSTFRQARVVGGTDADGEMPMQVSLHALQGHICGASLSPMLVSA 483  
 DB 317 GSDDEGGCS--RSSSPFLHRIVGSDSQDGTMPQVSLHFSVSA--CGASVISRMLLSAA 374  
 QY 484 HCYIDDRGFRTYDTQMTAFGLHDQSOR--APGQVRRILKRIISHPFNDFPFYDIA 541  
 DB 375 HCF---HGNRLSDPTPTAHLGMYVQGNAKFISF-----VRRIVHEVYNSQTFDYDIA 425  
 QY 542 LLELE--KPAEYSSAVRPICLPDASHVPAKAIIMVWGHT-QYCGGAILILKGEIRV 598  
 DB 426 LLDQSIAMPELTKQIDPICIPLPGQKVRSGKCVTGMGRHEDSGSVLQDAEVEL 485  
 QY 599 INQTTCENTLPOQITPRMVCVGLSGVDSCQDSCGGLS--SVEADGRIPQAGVVSMDG 657  
 DB 486 IDQTVVCSTY-GITTSRLCAGVWVGSKSDACKGDSGGFLSRRKRSDDGMILTGIVSWHG 544

QY 658 CAQNRKPYTRLPLFRDWI 677  
 DB 545 CGRNPFGVITRVSSFPVWI 564  
 RESULT 12  
 ID Q6PF94 PRELIMINARY; PRT; 799 AA.  
 AC Q6PF94  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 OS Trypsin-like protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=FVB/N; TISSUE=liver;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshlyuk S., Caraminci P., Prange C.,  
 RA Rana S.S., Loguellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McGwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Nadeau A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield J.S.,  
 RA Krzyzanski M.I., Skaleka U., Smalins D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=FVB/N; TISSUE=liver;  
 RL Strauberg R.;  
 CC Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to peptidase family S1.  
 DR EMBL: BC057674; AAH57674.1; --  
 DR GO; GO:0005886; C:plasma membrane; IDA.  
 DR InterPro: IPR000859; CUB.  
 DR InterPro: IPR002172; LDL\_receptor\_A.  
 DR InterPro: IPR001254; Peptidase\_S1.  
 DR InterPro: IPR001314; Peptidase\_S1A.  
 DR InterPro: IPR009003; Pept\_Ser\_Cys.  
 DR Pfam; PF00057; LDL\_recept\_a; 2.  
 DR Pfam; PF00089; Trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR SMART; SM00042; LDLa; 1.  
 DR SMART; SM00020; TRYP\_SPE; 1.  
 DR PROSITE; PS01180; CUB; 1.  
 DR PROSITE; PS01209; LDLRA\_1; 1.  
 DR PROSITE; PS50068; LDLRA\_2; 3.  
 DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 DR Hydrolase; Protease; Serine protease.  
 KW SEQUENCE 799 AA; 89557 MW; 16315A646AD5288 CRC64;  
 SQ  
 Query March 26.8%; Score 1014; DB 2; Length 799;  
 Best Local Similarity 34.4%; Pred. No. 7.2e-63;  
 Matches 213; Conservative 107; Mismatches 233; Indels 66; Gaps 15;



```

QY 72 COMALRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTYPSPSYNL 131
D 233 CLMHLQGPEDLMITKRL-EMTRVDCRDR-----VAMTDAAGPLEKRLITSYVGCGRQSPVM 287
QY 132 TFRSSQVLLITLITNTERRHNGF---EATFQPLRMS-SCGRLRKAQGTENSPYYPG 186
D 288 EVLASGVMAVVMKKGMHSYDPLLSVKSVAFOQCVNLTLEGR-L-DTQGFRTPTYPFS 346
QY 187 HYPENIDCTNIEVPPNNQHVVRKFFYLLEPRRACGTCPDYVEINGEKCYGER----- 241
D 347 YTSSTHCSMHLTPTSLDYLALMFDAYALRRQKYNRLCTQGGMMIQNRLCGRTTQPY 406
QY 242 SQFVVTNSNKTIVRFHSDSYDTDTGFLAEYLSYSSDPGQFTCRGTCIRKELRCDG 301
D 407 AERIPWASDGVITNFTSQISLTGPGVQVYVSLYNSQDPCRGFLGSV----- 454
QY 302 WADCTHSDSLNCSCDAGHQFTCKNFKCPLFWVCDSVNDGNSDEQCSCPAQTRC- 360
D 455 -----NGLCVP---ACDGIKDCPNGLDRNCVCGRAM-FOCQ 486
QY 361 SNGKCLSKSQCKGDKDCCGDSDEASCPKYNVYCTKHTRYCLNGCLSKSGNPECCKED 420
D 487 EDSTCISLPRVCDRQDPCLNGSDEQCOE--GVPCGFTTQCEPDRSCVCKRPNPECQSD 544
QY 421 CSDGSDKDCDGLRSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 480
D 545 CRDSDSEBHCDCGLQGLS--SRIVGTVSSEGMFWQASLIQIRGR-HICGALILADRMI 601
QY 481 SAACHYIDDRGFRYSPTQWTAFLGLHDQSORAPQVQERRLKRIISHPFNDFTFYDI 540
D 602 TAAHCFQED---SMASPKLMTVFLGKMRQNSR--WPEGVSFKVSRLFLHPHNEEDSHDYD 657
QY 541 ALLELEKPAEYSMWPRICLPDASHVFPAGKAIWVTGKHTQYCGTALILQKEIRVIN 600
D 658 ALLQDLHPVYTSATVPRVCLPARSHFFEPQHCWITGKAGRGEGPSNTLQKDVQVLP 717
QY 601 QTTCEMLLPQOITPRMNCVGLSGVDSCOGDSGGLPSVEADRIFOAGVWSMGDCAQ 660
D 718 QDLCSEAYRYQVSPRMLCAGYRKGRKDACQDSGGLVCRBPSGRWFLAGLVSMGLGCGR 777
QY 661 RNKRGVYTRLPFRDWKE 679
D 778 PNFFGVYTRVTRVINMIQ 796

```

```

RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchan J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzinski M.I., Skalka U., Smalins D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Matra M.A.;
RT "generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=liver;
RA Straubeberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC057674; AAH57674.1; -.
SQ SEQUENCE 799 AA; 89557 MW; 16315A646A4D5288 CRC64;

Query Match 26.8%; Score 1014; DB 2; Length 799;
Best Local Similarity 34.4%; Pred. No. 7,2e-63;
Matches 213; Conservative 107; Mismatches 233; Indels 66; Gaps 15;

QY 72 COMALRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTYPSPSYNL 131
D 233 CLMHLQGPEDLMITKRL-EMTRVDCRDR-----VAMTDAAGPLEKRLITSYVGCGRQSPVM 287
QY 132 TFRSSQVLLITLITNTERRHNGF---EATFQPLRMS-SCGRLRKAQGTENSPYYPG 186
D 288 EVLASGVMAVVMKKGMHSYDPLLSVKSVAFOQCVNLTLEGR-L-DTQGFRTPTYPFS 346
QY 187 HYPENIDCTNIEVPPNNQHVVRKFFYLLEPRRACGTCPDYVEINGEKCYGER----- 241
D 347 YTSSTHCSMHLTPTSLDYLALMFDAYALRRQKYNRLCTQGGMMIQNRLCGRTTQPY 406
QY 242 SQFVVTNSNKTIVRFHSDSYDTDTGFLAEYLSYSSDPGQFTCRGTCIRKELRCDG 301
D 407 AERIPWASDGVITNFTSQISLTGPGVQVYVSLYNSQDPCRGFLGSV----- 454
QY 302 WADCTHSDSLNCSCDAGHQFTCKNFKCPLFWVCDSVNDGNSDEQCSCPAQTRC- 360
D 455 -----NGLCVP---ACDGIKDCPNGLDRNCVCGRAM-FOCQ 486
QY 361 SNGKCLSKSQCKGDKDCCGDSDEASCPKYNVYCTKHTRYCLNGCLSKSGNPECCKED 420
D 487 EDSTCISLPRVCDRQDPCLNGSDEQCOE--GVPCGFTTQCEPDRSCVCKRPNPECQSD 544
QY 421 CSDGSDKDCDGLRSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 480
D 545 CRDSDSEBHCDCGLQGLS--SRIVGTVSSEGMFWQASLIQIRGR-HICGALILADRMI 601
QY 481 SAACHYIDDRGFRYSPTQWTAFLGLHDQSORAPQVQERRLKRIISHPFNDFTFYDI 540
D 602 TAAHCFQED---SMASPKLMTVFLGKMRQNSR--WPEGVSFKVSRLFLHPHNEEDSHDYD 657
QY 541 ALLELEKPAEYSMWPRICLPDASHVFPAGKAIWVTGKHTQYCGTALILQKEIRVIN 600
D 658 ALLQDLHPVYTSATVPRVCLPARSHFFEPQHCWITGKAGRGEGPSNTLQKDVQVLP 717
QY 601 QTTCEMLLPQOITPRMNCVGLSGVDSCOGDSGGLPSVEADRIFOAGVWSMGDCAQ 660
D 718 QDLCSEAYRYQVSPRMLCAGYRKGRKDACQDSGGLVCRBPSGRWFLAGLVSMGLGCGR 777
QY 661 RNKRGVYTRLPFRDWKE 679
D 778 PNFFGVYTRVTRVINMIQ 796

```

```

RESULT 14
TMS6 MOUSE STANDARD; PRT; 811 AA.
AC 09DBIO;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)

```

DE Transmembrane protease, serine 6 (EC 3.4.21.-) (Matrilinease-2).  
CN Name=Imp86;  
OS Mus musculus (mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN  
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi U., Bono H., Kondo S.,  
RA Nikiado I., Osato N., Saito R., Suzuki H., Yamana K.I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,  
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schirni L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Choitha C., Corbini L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimonard S., Guethrich S., Hitokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedziarski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lennard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagasima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sadelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Wanabe Y., Wells S.,  
RA Wilmberg L.G., Wynshaw-Boris A., Yangsong M., Yang L., Yang L.,  
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carlini P., Hayatsu N.,  
RA Hitzogane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
RA Shitaki T., Waki K., Kawai J., Aizawa K., Atakawa T., Fukuda S.,  
RA Hara A., Haeshizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2003).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=livrer;  
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schutler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bat N.G.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Ditschenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein W.T.J., Udell T.B., Toshiyuki S., Carinini P., Prange C.,  
RA Rana S.S., Loguoglu N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McGwan P.J., McKernan K.J., Malek J.A., Guarnarata P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Blakesley J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzyzanski M.I., Skalka U., Smallus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RE	[4]	REVIEW	
RP		MEDLINE=22668120; PubMed=12784999;	
RX		Netzel-Arnelt S., Hooper J.D., Szabo R., Madison E.L., Quigley J.P.,	
RA		Bugge T.H., Antalis T.M.;	
RA		"Membrane anchored serine proteases: a rapidly expanding group of cell	
RT		surface proteolytic enzymes with potential roles in cancer."	
RL		Cancer Metastasis Rev. 22:237-256(2003).	
CC		- FUNCTION: May play a specialized role in matrix remodeling	
CC		processes in liver (By similarity).	
CC		- SUBCELLULAR LOCATION: Type II membrane protein.	
CC		- TISSUE SPECIFICITY: Expressed at highest levels in liver, kidney	
CC		and uterus.	
CC		- DEVELOPMENTAL STAGE: Expressed at higher levels from 12.5 dpc to	
CC		15.5 dpc with a peak at 13.5 dpc. Expression in the developing	
CC		liver as well as a restricted set of embryonic epithelial cells of	
CC		the nasal cavity and pharyngo-lymphatic tubes.	
CC		- SIMILARITY: Belongs to peptidase family S1.	
CC		- SIMILARITY: Contains 2 CUB domains.	
CC		- SIMILARITY: Contains 3 LDI-receptor class A domains.	
CC		-----	
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC		the European Bioinformatics Institute. There are no restrictions on its	
CC		use by non-profit institutions as long as its content is in no way	
CC		modified and this statement is not removed. Usage by, and for commercial	
CC		entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC		or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC		-----	
DR		EMBL, AY240929; AAP69827.1; -	
DR		EMBL, AK004939; BAB23684.2; -	
DR		EMBL, BC029645; AAH29645.2; -	
DR		HSSP, P00760; IEXX.	
DR		MEROPS, S01.308; -.	
DR		MGD, MGI:191903; Tmpres6.	
DR		GO, GO:0016021; C:integral to membrane; ISS.	
DR		GO, GO:0005886; C:plasma membrane; IDA.	
DR		GO, GO:0004252; F:serine-type endopeptidase activity; ISS.	
DR		GO, GO:0001555; P:angiogenesis; ISS.	
DR		GO, GO:0030189; P:extracellular matrix organization and bioge. .; ISS.	
DR		GO, GO:0042730; P:fibrinolysis; ISS.	
DR		GO, GO:0007242; P:intracellular signaling cascade; ISS.	
DR		GO, GO:0006508; P:proteolysis and peptidolysis; ISS.	
DR		InterPro, IPR000859; CUB.	
DR		InterPro, IPR002172; LDI_receptor A.	
DR		InterPro, IPR001254; Peptidase_S1.	
DR		InterPro, IPR001314; Peptidase_S1A.	
DR		InterPro, IPR009003; Pept_Ser_Cys.	
DR		Pfam, PF00057; LDI_recept_a; 2.	
DR		Pfam, PF00089; TRYPSIN_1.	
DR		PRINTS, PR00722; CHYMOTRYPSIN.	
DR		PRINTS, PR00261; LDI-RECEPTOR.	
DR		PROSITE, PS01180; CUB: 1	
DR		PROSITE, PS01209; LDI-RA_1; 1.	
DR		PROSITE, PS50068; LDI-RA_2; 3.	
DR		PROSITE, PS50240; TRYPSIN_DOM; 1.	
DR		PROSITE, PS00134; TRYPSIN_HIS; 1.	
DR		PROSITE, PS00135; TRYPSIN_SER; 1.	
KW		Glycoprotein, Hydrolase, Repeat, Serine protease, Signal-anchor;	
KW		Transmembrane.	
FT		DOMAIN 1 59	
FT		TRANSMEM 60 80	
FT			Cytoplasmic (Potential).
FT			Signal-anchor for type II membrane
FT			protein (Potential).
FT			Extracellular (Potential).
FT		DOMAIN 81 811	
FT		DOMAIN 213 336	
FT		DOMAIN 323 440	
FT		DOMAIN 445 477	
FT		DOMAIN 478 514	
FT		DOMAIN 518 555	
FT		DOMAIN 565 799	
FT		ACT_SITE 617 617	
FT		ACT_SITE 668 668	
FT		ACT_SITE 762 762	
FT			Charge relay system (By similarity).
FT			Charge relay system (By similarity).
FT			Charge relay system (By similarity).

```

FT CARBOHYD 138 138 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 184 184 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 216 216 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 338 338 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 433 433 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 453 453 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 518 518 N-linked (GlcNAc...) (Potential)
FT CONFLICT 690 690 P -> PP (in Ref. 2)
SQ SEQUENCE 811 AA; 90978 MW; 32EB3E7C3127801B CRC64;

Query Match 26.8%; Score 1014; DB 1; Length 811;
Best Local Similarity 34.4%; Pred. No. 7.3e-63;
Matches 213; Conservative 107; Mismatches 233; Indels 66; Gaps 15;

QY 72 COMALRADADSVLSLTFERSFLASCDGRSDLVVTVTLSPMEPHALVQCGTTPPSYNL 131
Db CWMHLOGPEDMIVRL-EWTRVDCRR-----VAMVDAAGLEGRLLTSVYGCGRQSPVM 299
QY 132 TPISSQNVLLTLTLTTRRHHPGF-----EATFQGLPMS-SCGRLKKAQGTENSPYPG 186
Db EVLASGSVMAVVMKKGMHSYDPEFLSVKSAVFQDCVNLTLEGRLL-DTQGLRTPPYPS 358
QY 187 HYPNIDCTNNIEVPNNQHKVRFKFTYLLERACCTCPEDVYEINGEKYGER----- 241
Db YSPSTHSCWHLTVPSLDYGLALWFDAVALRQKYNRLCTQGWMIONRRLCGFRLLQPY 418
QY 242 SDFVTSNKNITVRFHSDSYDTDTGFLAEYLSYSDSPCEGQFTCTGRICRKLACDG 301
Db AERIPWASDGVITNFTSQTSLTGPQVQVYSLYNSDPCRGELCSV----- 466
QY 302 WADCTHSDINLSCDAHQFTCKNFKCKPLFWVCDSDVNDGSDSDQSCCPAQTRC- 360
Db 467 -----NGLCVP-----ACDGIKDCPNGLDERNCVCGRAM-FCQ 498
QY 361 SNGKCLSKSQCKNDKCGDSDASCPKXNVVTCTHTYRCLNGLSLKGNPCKDKED 420
Db 499 EDSTCISLPRVCRQPCPLNGSDDEQCE--GVPCGFTTQCEBRSQVKKPNPECDQSD 556
QY 421 CSDSDKDCDCGSLFTRQARVVGTDADGEMVPMVSLHALGQGHICGSLSPMWLV 480
Db 557 CRDSDQHCHDCGLQGIS--SRIVGVSSBEGEPWQASLQIRK-HICGALITADNWI 613
QY 481 SAACHYIDRGFRYSDEPTQWTAFLGLHDQSORSAVQGEERLKRILSHPFNDFDYDI 540
Db 614 TAAACPFQED---SWASPKMTVFLGKKRQNSR-WPGEVSFVRSILFLPHYHEDSHDYD 669
QY 541 ALLELEKPAEYSSNVRPCLPDASHVVPAGKATVWTGHTQYGTGALLIQKEIRVIN 600
Db 670 ALIQLDHPVYVSATVRVCLPARSHFPEPQHCMTGTGQREGGPPVSNLTQKVDVLP 729
QY 601 QTTCEMLTPOQITPRMNCVGFSLSGVSDSCGDSGGLPSYEADGRIFQAGVSWDGCQAQ 660
Db 730 QDLCSEAYRYQVSRPMICAGYRRKKKADACQSDSGPLVCBPPSGRWFLAGLVSGLCGR 789
QY 661 RNKPGVYTRLPFLFDWIK 679
Db 790 PNFGVYTRVTRVINVNIQ 808

RESULT 15
TMS6 HUMAN STANDARD; PRT; 811 AA.
AC 081U80; 081U82; 081XV8;
DT 10-OCT-2003 (Ref. 42, Created)
DT 10-OCT-2003 (Ref. 42, Last sequence update)
DT 05-JUL-2004 (Ref. 44, Last annotation update)
DE Transmembrane protease, serine 6 (EC 3.4.21-) (Matrilinease-2).
GN Name=TMPRSS6;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Hooper J.D., Quigley J.P.;
RT "TMPRSS6, a new type II transmembrane serine protease.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBS databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208; DOI=10.1038/990031;
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
RA Clamp M., Sunk L.J., Alnough R., Almeida J.P., Babbage A.K.,
RA Bagshaw C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cooley V.B., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gburek J.G.R., Goward M.E., Graffham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Marlyn J.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA O'Neill C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsey H., Ranney Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Winstonsia S., Kawasaki K., Sasaki T., Asakawa S., Kodon J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malai E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shauli S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan Q., Zhang G., Chisoe S., Murray J., Miller N., Mux P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshe H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Olesky P., Rohlfing T.,
RA Scheer P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Koef I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurashiki H., Salta S.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Belman L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Seroussi E., Franssen I., Tapia I., Bruder C.B.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tiliannu Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [3]
RP SROUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schaller G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wafer J., Heish F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M., Soares M.B., Bonaldo W.F., Casavant T.L., Scheetz T.E.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abreson R.D., Mallary S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Schmecher A., Schein J.E., Jones S.J.M., Maitra W.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";

```

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

[4] SEQUENCE OF 9-811 FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.

RP TISSUE=Fetal liver;

RC MEDLINE=22241917; PubMed=12149247; DOI=10.1074/jbc.M203007200;

RA Velasco G., Cal S., Quesada V., Sanchez L.M., Lopez-Otin C.;

RT "Matriptase-2, a membrane-bound mosaic serine protease predominantly expressed in human liver and showing degrading activity against extracellular matrix proteins";

RT J. Biol. Chem. 277:37637-37646 (2002).

RN [5]

RP MEDLINE=22668120; PubMed=12784999;

RA Netzel-Arnett S., Hooper J.D., Szabo R., Madison E.L., Quigley J.P., Bugge T.H., Antalis T.M.;

RT "Membrane anchored serine proteases: a rapidly expanding group of cell surface proteolytic enzymes with potential roles in cancer";

RT Cancer Metastasis Rev. 22:237-258 (2003).

CC -1- FUNCTION: May play a specialized role in matrix remodeling processes in liver.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named Isoforms=2;

CC Name=1;

CC IsoId=Q8IU80-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q8IU80-2; Sequence=VSP\_008379, VSP\_008380;

CC Note=No experimental confirmation available;

CC -1- TISSUE SPECIFICITY: Liver specific.

CC -1- SIMILARITY: Belongs to peptidase family S1.

CC -1- SIMILARITY: Contains 2 CUB domains.

CC -1- SIMILARITY: Contains 3 LDL-receptor class A domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AJ19876; CAC85953.1; ALT\_INIT.

DR EMBL: AY055383; AAL16413.1; -

DR EMBL: AY055384; AAL16414.1; -

DR EMBL: AL022314; -; NOT\_ANNOTATED\_CDS.

DR EMBL: BC039082; AAH39082.1; -

DR HSSP: P00760; 1EXX.

DR Genew: HGNC:16517; TMPRSS6.

DR InterPro: IPR000859; CUB.

DR InterPro: IPR002172; LDL\_receptor\_A.

DR InterPro: IPR001254; Peptidase\_S1.

DR InterPro: IPR001314; Peptidase\_S1A.

DR InterPro: IPR009003; Pept\_Ser\_Cys.

DR Pfam: PF00431; CUB; 1.

DR Pfam: PF00057; LDL\_recept\_a; 2.

DR Pfam: PF00089; Trypsin\_1.

DR PRINTS: PR00722; CHYMOTRYPSIN.

DR PRINTS: PR00261; LDLRECEPTOR.

DR SMART: SM00042; CUB; 1.

DR SMART: SM00192; LDLa; 3.

DR SMART: SM00020; Tryp\_spc; 1.

DR PROSITE: PS01180; CUB; 1.

DR PROSITE: PS01209; LDLRA\_1; 1.

DR PROSITE: PS50068; LDLRA\_2; 3.

DR PROSITE: PS50240; TRYPSIN\_DOM; 1.

DR PROSITE: PS00134; TRYPSIN\_HIS; 1.

DR PROSITE: PS00135; TRYPSIN\_SER; 1.

KW Alternative splicing; Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor; Transmembrane.

FT DOMAIN 1 55 Cytoplasmic (Potential).

FT TRANSMEM 56 76 Signal-anchor for type II membrane protein (Potential).

FT DOMAIN 77 811 Extracellular (Potential).

FT DOMAIN 213 336 CUB 1.

FT DOMAIN 335 452 CUB 2.

FT DOMAIN 457 489 LDL-receptor class A 1.

FT DOMAIN 490 526 LDL-receptor class A 2.

FT DOMAIN 530 567 LDL-receptor class A 3.

FT DOMAIN 577 611 Serine protease.

FT ACT\_SITE 617 617 Charge relay system (By similarity).

FT ACT\_SITE 668 668 Charge relay system (By similarity).

FT ACT\_SITE 762 762 Charge relay system (By similarity).

FT CARBOHYD 136 136 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 184 184 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 216 216 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 338 338 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 433 433 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 453 453 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 518 518 N-linked (GlcNAc...) (Potential).

FT VARSPLIC 409 461 LCGRLIQPAERIPVAVATGITINFTSILSTPGVAVH GLYNOSDPCGE -> YHPLSSIMLPLPPPPSPSTVTP SLEAQVNLRGASRGSMQCACP (in isoform 2).

FT VARSPLIC /FTId=VSP\_008379.

FT VARSPLIC Missing (in isoform 2).

FT CONFLICT 116 116 /FTId=VSP\_008380.

FT SEQUENCE 811 AA; 8999 MM; 78EP193P65SDSD CR64;

Query Match 26.6%; Score 1007; DB 1; Length 811;

Best Local Similarity 34.0%; Pred. No. 2,3e-62;

Matches 217; Conservative 99; Mismatches 229; Indels 94; Gaps 17;

67 PAH--ARCOMALRGDADSVSLTFRSFDLASCDERGSGLTVYVNTTSPMEPHALVOL--C 122

238 PDHILASCLMHLQCPKDMKLRL-EWTLAECHDR---LAMDVAGPLKRLITLTSYGC 292

123 GYPPPSYNTL-----FHSSQNVLLTLTNTERRRHDFEATFQGLPRMSSCG 169

293 SRQEPVVEVLASGAIMAVVWKKGLHSYDDPFLSV-----QPVVFOACEVALTL 341

170 GLRLKAGCTNSPPYRPHYRPNDCTMINIEVPNNQHYKRFKPYLLERRAGCTPKDY 229

342 DNRLDGQGLSTPYFPSPYSPQTHCSWHLTVPSLDYGLALMFAYALRRKXYDLPTQGG 401

230 VEINGEKYCG-----ERSQFVTSNSNKKITVRFHSDSYTDYGLAEYLSYDSDPC 281

402 WTLQNRLLCGRLIQPAERIPVAVATG---ITINFTSILSTPGVAVHGLYNOSDPC 458

282 PGQFTCRTRGCIKRELCDGMADCTDHSDELNCSDAHQFTCKNKKCPFLFWVCSVND 341

459 GGEFLGCV-----NGLCVF---ACDGVKD 479

342 CGDNSDEQSCSPAQTFRC-SNGKLSKSQQCKGDKDGGSDASCPKXNVVYCTKHTY 400

480 CPNGLDERNCVCRA-FTQCKEDSTCISLTPVCGQDPCLNGSDEEQOE--GVPCGTFE 536

401 RCLNGLCTSKGNPCDCKEDPCDSDGDEKDCDGLRSTROARVAVGGDADAGEPMQVSL 460

537 QCEDRSCTKKPNPCDCKDPCDSDGDEHCDGLG--ESSRIYGVGAVSSEGEPMQASL 594

461 HALGQGHICGASLSPMVLVSAHCYIDRGFPYSPDTQWTAFILGLHDSQASAPGOER 530

595 QVNGR-HICGGLIADRWVITAHCQED---SWASTVLTWTFVLGKWKQNSR-WPGEVSF 649

521 RLKRIISHPPNDFTFYDIALLEKPAEYSSMWKPICLPDASHVTPAKAIVTWGMH 580

650 KVSRLLAHPHEBDSHYDVALQLDHPVVRSAVAVVCLPARSHFEPELHCMTWGA 709

581 TOYGGTALILQGEIRIVNQTCCNLLPQGITPRMVCVFLSGVDSCGSDGSPLSV 640

710 LRESGPLSNALQKXVDVLIQDLCSEVYRYQVTPRMLCAGYRGKGDACGSDGGPLVCK 769

641 EADGRIFQAGVSWGDCAQRNRPVYTRLLPLFRDWIKE 679

770 ALSGRWFLAGLVSMGLGCGRPYFVGYTRITGVISWIOQ 808

```
RESULT 16
Q6UXD8 PRELIMINARY; PRT; 802 AA.
ID Q6UXD8;
AC Q6UXD8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE PVAE354.
GN ORFNames=UNO354;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22887296; PubMed=12975309;
RA Clark H.F., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,
RA Chen J., Chow B., Chui C., Crowley C., Currell B., Deuel B., Dowd P.,
RA Eaton D., Foster J., Grimaldi C., Gu Q., Hass P.B., Heldens S.,
RA Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,
RA Lewis L., Liao D., Mark M., Robbie E., Sanchez C., Schoenfeld J.,
RA Sehgiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,
RA Vandlen R., Watanabe C., Wleand D., Woods K., Xie M.H., Yansura D.,
RA Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,
RA Godowski P.;
RT "The secreted protein discovery initiative (SPDI), a large-scale
RT effort to identify novel human secreted and transmembrane proteins: a
RT bioinformatics assessment.";
RT Genome Res. 13:2265-2270(2003).
CC -1- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; AY358398; AAQ8764.1; -.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00057; Ldl_recept_a; 2.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 3.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LDLR_1; 1.
DR PROSITE; PS50068; LDLR_2; 3.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KM Hydrolase; Protease; Serine protease.
SQ SEQUENCE 802 AA; 88845 MW; 8726C91B1E02E163 CRC64;

Query Match 26.5%; Score 1005; DB 2; Length 802;
Best Local Similarity 34.0%; Pred. No. 3.1e-62;
Matches 217; Conservative 98; Mismatches 230; Indels 94; Gaps 17;
```

```
QY 282 PGQFTCRGCIKREIKCDGMADCTDHSDELNCSCDAGHQTCKNFKCPLFMVCDVSVD 341
DB 450 RGEFLCSV-----NGLCVP---ACGVDK 470
QY 342 CGNSDBQCGCPAQTFRRC-SNGKCLSKSQCGKDDCGDSDPBAACCPKVVYTCRHTY 400
DB 471 CPNGLBERNCVCA--PFQCKEDSTCISLPVCGQDCLNGSEBQCE--GVPCTFTF 527
QY 401 RCNGLCLSKNPECKEKEDCSGDSKDCDGLRSTFQARVVGTDDEGEMPVSL 460
DB 528 QCEDRSCVKKPNQCGRPDRCRDSBEHCDCQ--PSSRVGGAVSSEGMPQASL 565
QY 461 HALGQGHICASLISPNMIVSAACVYIDRGFRYSPTQWTAFLGHDQSQRSAFQVER 520
DB 586 QVRGR-HICGALILADRWVITAAHCFQED---SMASTVLTWVFLGKVMQNSR-WPGEVSP 640
QY 521 RLKRIISHPPENDFTDYDIALLEKPAEYSSMVRPICLPDASHPPAKAIWYGMGH 560
DB 641 KVRRLIHPYHEBDSHDYDVALQLDHPVRSAAVPCLPARSHFEPQLHCWITGMGA 700
QY 581 TOYGTGALILQGEIRVINTQTCENLTPQQTFRPMVCVGLSGVDSQGDGGLSSV 640
DB 701 LRGGPISNALQKVYDQLIPDLCSAARYQVTPRHLCAGRKXKDAQGDGSGFLVCK 760
QY 641 EADGRIFQAGVSWDGCAGQNRKPGVYTRLPFRDWIKE 679
DB 761 ALSGRWFLAGLVSMGLGCGRPNYFGVYTRITGVISMIQ 799
```

```
RESULT 17
AAQ88764 PRELIMINARY; PRT; 802 AA.
ID AAQ88764;
AC AAQ88764;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE PVAE354.
GN UNO354.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primata; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12975309;
RA Clark H.F., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,
RA Chen J., Chow B., Chui C., Crowley C., Currell B., Deuel B., Dowd P.,
RA Eaton D., Foster J., Grimaldi C., Gu Q., Hass P.B., Heldens S.,
RA Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,
RA Lewis L., Liao D., Mark M., Robbie E., Sanchez C., Schoenfeld J.,
RA Sehgiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,
RA Vandlen R., Watanabe C., Wleand D., Woods K., Xie M.H., Yansura D.,
RA Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,
RA Godowski P.;
RT "The Secreted Protein Discovery Initiative (SPDI), a Large-Scale
RT Effort to Identify Novel Human Secreted and Transmembrane Proteins: A
RT Bioinformatics Assessment.";
RT Genome Res. 13:2265-2270(2003).
DR EMBL; AY358398; AAQ88764.1; -.
SQ SEQUENCE 802 AA; 88845 MW; 8726C91B1E02E163 CRC64;

Query Match 26.5%; Score 1005; DB 2; Length 802;
Best Local Similarity 34.0%; Pred. No. 3.1e-62;
Matches 217; Conservative 98; Mismatches 230; Indels 94; Gaps 17;
```

```

Qy 170 GLRKAQGTNSPYRGHYPPNIDCTWNIIEVNNQHVKVRKFFYLLEPRRACGTCPKDY 229
Db 333 DNRLDSDQVLSLTPYFSTYSPTHTCSWHLTVPSLDYGLALFADYALRRQKIDLPCTGQ 392
Qy 230 VEINGEKYCG-----ERSQFVVTNSNKTIVRFHSDOSYTDGTGLAEYLSYDSDPC 281
Db 393 WTIGNRRLCGRLIQPYAERIIPVATAG---ITINFTSQISLTGPGVAVHGLVNGSDPC 449
Qy 282 PGQFTCTGRCIRKELRCDCMADCTDHSDELNCSCDAGHQTCKNKKCKPLFWVCDVND 341
Db 450 PGEFLCSV-----NGLCVP---ACDGVKD 470
Qy 342 CGDSDSDQSGCPAQTRFC-SNGKCLSKSQCNKDKDGDGSDASCPKVVVCTGHTY 400
Db 471 CPNGLDERNCVCRA-TFOCKEDSTISLPKVCDDQPDCLNSDEBQOE--GVPCGTFTF 527
Qy 401 RCLNGLCLSKNPECDKEDSDGSDKDCDGLRSTFROARVVGTDADGEMPMQVSL 460
Db 528 QCEBRSQVKNPCDGRPDGRSDSDHCHDGLGQ--PSRIVGAVSSGEMPMQVSL 585
Qy 461 HALQGHICGASLISPNVLVAHACYIDRGFRISDPTQWTAFLGLHDQSORAPVOER 520
Db 586 QVRGR-HICGGLALADRNVITPAHCFQED---SMASVTLMVTFGLKQVQNSR-WPGEVSF 640
Qy 521 RLKRIISHPFNDPTFDYDIALLEKPAEYSSWVRPCLPDASHVFPAGKAIWVTMGH 580
Db 641 KVSRLLPHPHEEDSHDYVALQLDHPVRSAAVRPCLPARSHFEPGLHCHWTMGGA 700
Qy 581 TOYGGTGA-LIIOKGEIRVINOTTCENLLPQOITPRPMCVGFLSGVDSQCGDGGPLSSV 640
Db 701 LRBGGPISNALQKDVQOLIPODLCEAVRYQVTPRMLCAGRKKKDACCQDSSGSLVCK 760
Qy 641 EADGRIFQAGVSVWGDGCAQRNKGVTYRLPLFRDWIK 679
Db 761 ALSGRMFLAGLWSGLGCGRPYFGVYTRITGVISMIQ 799

RESULT 18
ID Q6IC2 PRELIMINARY; PRT; 824 AA.
AC Q6IC2;
DT 05-JUN-2004 (TREMBlrel. 27, Created)
DT 05-JUN-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUN-2004 (TREMBlrel. 27, Last annotation update)
DE D1170K4.2 protein.
GN Name=d1170K4.2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,
RA Cole C.G., Goward M.E., Aguado B., Mally M., Mokrab Y., Huckle E.J.,
RA Beare D.M., Dunham I.,
RA Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family S1.
EMBL; CR456446; CAG30332.1; -.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1.
DR InterPro: IPR009003; Pept_Ser_Cys.
DR Pfam: PF00431; CUB; 1.
DR Pfam: PF00057; Ldl1_recept_a; 2.
DR Pfam: PF00089; Trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 3.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LDLRA_1; 1.

```

```

DR PROSITE; PS0068; LDLRA_2; 3.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydroxylase; Protease; Serine protease.
SQ SEQUENCE 824 AA; 9133 MW; A74F186406041F7B CRC64;

Query Match 26.0%; Score 985; DB 2; Length 824;
Best Local Similarity 32.8%; Pred. No. 8.2e-61;
Matches 217; Conservative 98; Mismatches 230; Indels 116; Gaps 18;

Qy 67 PAH--ARCOMALRGDADSVLSLTPRSFDLASCDERGSGLVTVVNTLSPEMEHALVOL--C 122
Db 229 PDHLASCLMHLQCPKMLKRLR-EWTLAEGR---LAMYVAGLEKRLTLTSYGC 283
Qy 123 GTYPPSTNLT-----FHSSONVLLTLINTERRHPGEATFPQLFRMSSCG 169
Db 284 SRQPVVVEVLASGAIMVWVKKGLHSYYDPVLSV-----QPVVFOACEVNLTL 332
Qy 170 GLRKAQGTNSPYRGHYPPNIDCTWNIIEVNNQHVKVRKFFYLLEPRRACGTCPKDY 229
Db 333 DNRLDSDQVLSLTPYFSTYSPTHTCSWHLTVPSLDYGLALFADYALRRQKIDLPCTGQ 392
Qy 230 VEINGEKYCG-----ERSQFVVTNSNKTIVRFHSDOSYTDGTGLAEYLSYDSDPC 281
Db 393 WTIGNRRLCGRLIQPYAERIIPVATAG---ITINFTSQISLTGPGVAVHGLVNGSDPC 449
Qy 282 PGQFTCTGRCIRKELRCDCMADCTDHSDELNCSCDAGHQTCKNKKCKPLFWVCDVND 341
Db 450 PGEFLCSV-----NGLCVP---ACDGVKD 470
Qy 342 CGDSDSDQSGCPAQTRFC-SNGKCLSKSQCNKDKDGDGSDASCPKVVVCTGHTY 400
Db 471 CPNGLDERNCVCRA-TFOCKEDSTISLPKVCDDQPDCLNSDEBQOE--GVPCGTFTF 527
Qy 401 RCLNGLCLSKNPECDKEDSDGSDKDCDGLRSTFROARVVGTDADGEMPMQVSL 460
Db 528 QCEBRSQVKNPCDGRPDGRSDSDHCHDGLGQ--PSRIVGAVSSGEMPMQVSL 585
Qy 461 HALQGHICGASLISPNVLVAHACYIDRGFRISDPTQWTAFLGLHDQSORAPVOER 520
Db 586 QVRGR-HICGGLALADRNVITPAHCFQED---SMASVTLMVTFGLKQVQNSR-WPGEVSF 640
Qy 521 RLKRIISHPFNDPTFDYDIALLEKPAEYSSWVRPCLPDASHVFPAGKAIWVTMGH 580
Db 641 KVSRLLPHPHEEDSHDYVALQLDHPVRSAAVRPCLPARSHFEPGLHCHWTMGGA 700
Qy 581 TOYGGTGA-----LIIOKGEIRVINOTTCENLLPQOITPRPMCVGFLSGVDSQCGDGGPLSSV 618
Db 701 LRBGALADAVALFYGRNQGSEITCCPISNALQKDVQOLIPODLCEAVRYQVTPRMLC 760
Qy 619 VGLSGVDSQCGDSSGFLSVEADGRIFOAGVSVWGDGCAQRNKGVTYRLPLFRDWIK 678
Db 761 AGYKKGKDKACQDSSGFLVCKALSGRMFLAGLWSGLGCGRPYFGVYTRITGVISMIQ 820
Qy 679 E 679
Db 821 Q 821

RESULT 19
ID CAG30332 PRELIMINARY; PRT; 824 AA.
AC CAG30332;
DT 01-JUN-2004 (TREMBlrel. 27, Created)
DT 01-JUN-2004 (TREMBlrel. 27, Last sequence update)
DT 01-JUN-2004 (TREMBlrel. 27, Last annotation update)
DE D1170K4.2 protein.
GN D1170K4.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;

```

```

RN [1]
SEQUENCE FROM N.A.
RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,
RA Cole C.G., Goward M.E., Aguado B., Mallya M., Mokrab Y., Huckle E.J.,
RA Beare D.M., Dunham I.
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, CR456446; CA630332.1; -.
SQ SEQUENCE 824 AA; 91333 MW; A74F186406041F7B CRC64;

Query Match 26.0%; Score 985; DB 2; Length 824;
Best local Similarity 37.8%; Pred. No. 8.2e-61;
Matches 217; Conservative 98; Mismatches 230; Indels 116; Gaps 18;

QY 67 PAH--ARCOVALRGDADSVSLTRFSFDLASCDERGSGLVTVNTLSMPMEHALVQD--C 122
DB 229 PDHLASCLMHLQPKDLMRLRL--EWTLAECRDR---LAMYDVAPLEKRLITSYGC 283
QY 123 GTYPPSTNLT-----FHSSONVLLTLINTERRRHGFATPQLPRMSSCG 169
DB 284 SROEPVEVLASGAIMAVWKKGLHSYDDPFLSV-----OPVVFQACEVNLTL 332
QY 170 GRLKAQGTFSPPYPGHYPPNIDCTWNIIEVPNNQHYKVPKFFYLEPRRACGTCPKDY 229
DB 333 DNRLLDSQGLSTPYFPSTYSPTQTHCSMLTVPSLDYGIALMFDYALRRQKTYDLPTQCG 392
QY 230 VEINGEKYCG-----ERSQFVVTNSNKKITVRFHSDSYDPTGFLAEYLSYSDSPC 281
DB 393 WTIGNRRLCGRLIPYAEIRIPVAVATG---ITINFTSQSLTGPVAVHGLYNQSDPC 449
QY 282 PGQCTCTGTCIRKELACDGMADCTDHSDELNCSGDMGHQFTCKNKKCKPLFWCDSVND 341
DB 450 PGEFLCSV-----NGLCVP---ACDGVKD 470
QY 342 CGNDSDDQSCSPAQTFRC--SNGKCLSKSQCGKNGDGDGSDPASCPRKVVVTCTGHTY 400
DB 471 CPNGLDERNCVCRA--TQCKEDSTCTISLKYCDQPPCLNGBDEBQOE--GVPCGTFP 527
QY 401 RCLNGCLSKNPECDKEDCDSDSKDCDGLRSTFRQARVVGTDADGEMVQVSL 460
DB 528 QCEDRSQVKKPNCQDGRPCDRDSDREHCCGLOG--PSSRIYGAIVSSGEMVQVSL 585
QY 461 HALGQGHICGASLSPWLVSAACIYDDKGRFRSDPTQMTATPLGLHDGSRASPGYQR 520
DB 586 QVRER--HICGGLADRWVITPAACFQED--SNASTVLMVTFVFGKWNQSR--WEGVSP 640
QY 521 RLKRIISHPEFNDFTFYDIALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGMH 580
DB 641 KVSRLLLHPHEBDSHYDVALLDLHPVRSAAVRVCLPARSHFEPGLHCHITWGA 700
QY 581 TQYGGTGA-----LIIQGEIRVYNQTTCENTLLPQGITPRMMC 618
DB 701 LREGALRADAVALEFYGWRNQGSETCCCPISNALQKVAVQILPOLCSEVYRYQTTPMLC 760
QY 619 VGFPLSGVDSCQGGSGGPLSVSEADGRIFQAGVYSWMDGCAQRKRPVYTRLPFRMWIK 678
DB 761 AGYRKGGKACQGGSGGPLVCKALSGRWFLAGLVSWGLGCGRPVYFGVYTRITGVISMIQ 820
QY 679 E 679
DB 821 Q 821

RESULT 20
ENTK_PIG STANDARD; PRT; 1034 AA.
AC P98074;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
GN Name=PRSS7; Synonyms=ENTK;
OS Sus scrofa (Pig);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN NCBI_TaxID=9823;
RX [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Duodenal mucosa;
RX MEDLINE=94327548; PubMed=8051081;
RA Matsushima M., Ichinose M., Yahagi N., Kakei N., Tsukada S., Miki K.,
RA Kurokawa K., Tashiro K., Shiohara K., Shinomiya K., Uemeyama H.,
RA Inoue H., Takahashi T., Takahashi K.;
RT "Structural characterization of porcine enteropeptidase.";
RL J. Biol. Chem. 269:19976-19982(1994).
CC -1- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
CC trypsinogen.
CC -1- SUBUNIT: Heterotrimer of a catalytic (light) chain, a multidomain
CC (heavy) chain, and a mini chain.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- PTM: The chains are derived from a single precursor that is
CC cleaved by a trypsin-like protease.
CC -1- PTM: The mini chain may be cleaved by elastase.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 MAM domain.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, D30799; BA06459.1; -.
DR HSSP, P98072; 1EBX.
DR MR08S, S01.156; -.
DR InterPro, IPR000859; CUB.
DR InterPro, IPR002172; LDL_receptor_A.
DR InterPro, IPR000998; MAM.
DR InterPro, IPR001254; Peptidase S1.
DR InterPro, IPR001314; Peptidase S1A.
DR InterPro, IPR011163; Pept_S1A_enterop.
DR InterPro, IPR009003; Pept_Ser_Cys.
DR InterPro, IPR001190; Strc_receptor.
DR Pfam, PF00431; CUB; 2.
DR Pfam, PF00057; LDL_recept_a; 2.
DR Pfam, PF00629; MAM; 1.
DR Pfam, PF01350; SEA; 1.
DR Pfam, PF00530; SRCR; 1.
DR Pfam, PF00089; Trypsin; 1.
DR PIRSF, PIRSF001138; Enteropeptidase; 1.
DR PRINTS, PR00722; CHYMOTRYPSIN.
DR PRINTS, PR00261; LDLRECEPTOR.
DR PRINTS, PR00020; MAMDOMAIN.
DR SMART, SM00042; CUB; 2.
DR SMART, SM00192; LDLA; 2.
DR SMART, SM00137; MAM; 1.
DR SMART, SM00200; SEA; 1.
DR SMART, SM00202; SR; 1.
DR SMART, SM00020; TRY_SPC; 1.
DR PROSITE, PS01180; CUB; 2.
DR PROSITE, PS01209; LDLA_1; 2.
DR PROSITE, PS50068; LDLA_2; 2.
DR PROSITE, PS00740; MAM_1; 1.
DR PROSITE, PS50060; MAM_2; 1.
DR PROSITE, PS50024; SEA; 1.

```



```

DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Direct protein sequencing; Glycoprotein; Hydrolyase; Lipoprotein;
KW Myelastate; Repeat; Serine protease; Signal-anchor; Transmembrane;
KW Zymogen.
FT CHAIN 52 117 Enteroproteidase non-catalytic mini chain.
FT CHAIN 118 799 Enteroproteidase non-catalytic heavy
FT CHAIN chain.
FT CHAIN 800 1034 Enteroproteidase catalytic light chain.
FT DOMAIN 1 18 Cytoplasmic (Potential).
FT TRANSMEM 19 47 Signal-anchor for type II membrane
FT protein (Potential).
FT extracellular (Potential).
FT SEA.
FT LDL-receptor class A 1.
FT LDL-receptor class A 1.
FT CUB 1.
FT MAM.
FT CUB 2.
FT LDL-receptor class A 2.
FT SRCR.
FT Serine protease.
FT Charge relay system (By similarity).
FT ACT_SITE 840 840 Charge relay system (By similarity).
FT ACT_SITE 891 891 Charge relay system (By similarity).
FT ACT_SITE 986 986 Charge relay system (By similarity).
FT LITID 2 2 N-myristoyl glycine (Potential).
FT DISULFID 199 212 By similarity.
FT DISULFID 206 225 By similarity.
FT DISULFID 219 236 By similarity.
FT DISULFID 658 670 By similarity.
FT DISULFID 665 683 By similarity.
FT DISULFID 677 692 By similarity.
FT DISULFID 787 911 Interchain (By similarity).
FT DISULFID 825 841 By similarity.
FT DISULFID 925 992 By similarity.
FT DISULFID 956 971 By similarity.
FT DISULFID 982 1010 By similarity.
FT CARBOHYD 116 116 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 147 147 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 170 170 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 194 194 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 283 283 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 343 343 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 350 350 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 403 403 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 455 455 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 485 485 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 518 518 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 549 549 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 645 645 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 697 697 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 701 701 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 721 721 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 740 740 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 761 761 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 804 804 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 863 863 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 902 902 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 964 964 N-linked (GlcNAc... ) (Potential).
SQ SEQUENCE 1034 AA; 114776 MW; 0386C64CF6CC368 CRC64;

Query Match 19.0%; Score 718.5; DB 1; Length 1034;
Best Local Similarity 32.4%; Pred. No. 6.1e-42;
Matches 181; Conservative 87; Mismatches 189; Indels 101; Gaps 22;

```

```

Db 591 -----IADVBEIRDEEDSDLLAVYTGPGVEDVFTTNMTVLFTINDALTKGFK 643
QY 270 AEVL$---YDSOPC-RGQPTCRGRCIRKELRCQDGMADCTDHSDELNGSCDAGHQTCK 325
Db 644 ANFTTGYHLGIPECKEDNFOCENGECVLLVNLCDGSHCKDSDEAHCV----- 693
QY 326 NKPEKPLFWVCDSDVNDGSDGSCSPAQTRFCS---NGKLSK9QCGKXDC---G 379
Db 694 -RF-----LNGRANNSG-----LVGRIGSIWHTACAEWTTQTSDVQCGLG 735
QY 380 DGSDEASCP-----KVVVYTCYKHTYRCLNGLCLSKNPECKXED-----CSD 423
Db 736 LGTENSMPFSSGSGGPFVLTNP-----NGSLITLASBQC--FEDSLILQCN 784
QY 424 GSDKDDDCGLRSTFRA--RVVGCTDADGEMPMQVSLHALGQHTCGASLSPMLVS 481
Db 785 KS-----CGKQVQSVSPRIVGNSRSGAMPMVVALYYNGQ-LICGSLVSRDLVS 837
QY 482 AAHCYIDRGRFRYSDPTQWTAFLGLHDQSORSAFQVGRRLKRIISHPFNDPTFDYDIA 541
Db 838 AAHCYVG---RNLPEKMKAILGLHMTSNLTSFQVTRLIDELIVINPHYNRKXSDIA 893
QY 542 LLELEKPAEYSSWVRPILCPDASHVFPAGAIWVGHTQYGTGALLIQGEIRVINO 601
Db 894 MHLFEFVNTYDYOPICLPEENQVFPGRICSIAGKVIYQSSPADIIQEDVPLLSN 953
QY 602 TTGCENLDPQ-QIFRMMVCVFLSGGVNDCGDSGSPISVADGRIRPQGVNMGDCAQ 660
Db 954 EKCQOQPEVNTIENMMCAQIEBGGIDSCGDSGGLMCLN-NRMILAGVTSFGYCAL 1012
QY 661 RNKPGVYTRPLPFRDWIK 678
Db 1013 PNRGVYARVPKFTWMIQ 1030

RESULT 21
Q72410 PRELIMINARY; PRT; 855 AA.
ID 072410:
AC 072410:
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyease-1B protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Liver;
RX MEDLINE=22784742; PubMed=12886014;
RA Cai S., Lopez-Otin C.;
RT "Polyeasase-1, a human polypeptidase with the ability to generate
RT independent serine protease domains from a single translation
RT product.";
RL EMBL; AJ488947; CAD35759.1; -.
DR MEROPS; S01.357; -.
DR MEROPS; S01.358; -.
DR MEROPS; S01.969; -.
DR GO; GO:0004295; F:Trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00087; Ldl_recept_a; 1.
DR Pfam; PF00089; Trypsin_2.
DR SMART; SM00192; Ldla; 1.
DR SMART; SM00020; Tryp_SPC; 2.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 2.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_2.
DR PROSITE; PS00135; TRYPSIN_SER; 2.
KW Hydrolyase; Protease; Serine protease.

```

Seq	SEQUENCE	855 AA;	91352 MW;	8A22759D9740CF3F CRC64;
Query Match		18.9%;	Score 717.5;	DB 2;
Best Local Similarity		45.7%;	Pred. No. 5.8e-42;	
Matches 132;		Conservative 53;	Mismatches 95;	Indels 9;
Gaps				5
Qy	CTKHTRYCLNGLCLSGKGNPECGKDKDSCDSDPEXDCDGLSFTFQA-RVVGSTADDEG			453
Db	154 CPGNFSFGNSGCVTKVKNPECDQDQDCSDGSEBAHCECGIQAMRWAGHIVGMEASPEE			213
Qy	454 WPMOVSIHALQGHICGASLISPNMLVSAAMCYIDDGRFRSDPTQWTAFLGLIHDSQRS			513
Db	214 FPMQASLIRE-NKEHFCGAIIINARWLVSAAHCENE-----FQDPKRWAVYAGATYLSGSE			267
Qy	514 APQGOERLKLIIHPFNDFPFVDIALLLEKPAEYSMWKRPICLPASAVFPAGKAI			573
Db	268 ASTVR-AQVVOIVKRPILYNADPADPADVAVLELISPLPFGRIHQVCLPAPATHIFPSSKC			326
Qy	574 WVTGNGHYQYGG-TGALLLOKEIIVINOTTENILLPQITPRMNCVGLSGVNSCGD			632
Db	327 LISNGYLIKEDPLVKEPVYQKRTVELLDALCASLYGHSLTDRMVCAGLIDKVDSCGD			386
Qy	633 SGGPLSSVEADGRILFQAGVSWGDCCAQRMKPGVYTRLPFLFDWKENT			681
Db	387 SGGPLVCEPSGSRFPLAGIVSWGICGAERRRGVAVTRLDWILIENT			435
RESULT 22				
ID	Q72411			
AC	Q72411	PRELIMINARY;	PRT;	1059 AA.
DT	01-OCT-2003	(TREMBLrel. 25, Created)		
DT	01-OCT-2003	(TREMBLrel. 25, Last sequence update)		
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)		
DS	Polyserase-1A protein.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=22784742; PubMed=12886014;			
RA	Cal S., Lopez-Otin C.;			
RT	"Polyserase-1, a human polypeptidase with the ability to generate			
RT	independent serine protease domains from a single translation			
RT	product."			
RL	Proc. Natl. Acad. Sci. U.S.A. 100:9185-9190(2003).			
DR	EMBL; AJ488946; CAD35758.1; --.			
DR	MEROPS; S01.357; --.			
DR	MEROPS; S01.358; --.			
DR	MEROPS; S01.969; --.			
DR	GO; GO:0004295; F:trypsin activity; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro; IPR002172; LDI_Receptor_A.			
DR	InterPro; IPR001254; Peptidase_S1.			
DR	Pfam; PF00057; LDI_recept_a; 1.			
DR	Pfam; PF00089; Trypsin_3.			
DR	SMART; SM00192; LDIa; 1.			
DR	SMART; SM00020; Tryp_SPC; 3.			
DR	PROSITE; PS50068; LDIa_2; 1.			
DR	PROSITE; PS50240; TRYPSIN_DOM; 3.			
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_3.			
DR	PROSITE; PS00135; TRYPSIN_SER; 2.			
KM	Hydrolase; Protease; Serine protease.			
SO	SEQUENCE	1059 AA;	114020 MW;	17D27A2D992A264 CRC64;
Query Match		18.9%;	Score 717.5;	DB 2;
Best Local Similarity		45.7%;	Pred. No. 7.3e-42;	
Matches 132;		Conservative 53;	Mismatches 95;	Indels 9;
Gaps				5
Qy	CTKHTRYCLNGLCLSGKGNPECDGKDKDSCDSDPEXDCDGLSFTFQA-RVVGSTADDEG			453
Db	154 CPGNFSFGNSGCVTKVKNPECDQDQDCSDGSEBAHCECGIQAMRWAGHIVGMEASPEE			213
Qy	454 WPMOVSIHALQGHICGASLISPNMLVSAAMCYIDDGRFRSDPTQWTAFLGLIHDSQRS			513
Db	214 FPMQASLIRE-NKEHFCGAIIINARWLVSAAHCENE-----FQDPKRWAVYAGATYLSGSE			267
Qy	514 APQGOERLKLIIHPFNDFPFVDIALLLEKPAEYSMWKRPICLPASAVFPAGKAI			573
Db	268 ASTVR-AQVVOIVKRPILYNADPADPADVAVLELISPLPFGRIHQVCLPAPATHIFPSSKC			326
Qy	574 WVTGNGHYQYGG-TGALLLOKEIIVINOTTENILLPQITPRMNCVGLSGVNSCGD			632
Db	327 LISNGYLIKEDPLVKEPVYQKRTVELLDALCASLYGHSLTDRMVCAGLIDKVDSCGD			386
Qy	633 SGGPLSSVEADGRILFQAGVSWGDCCAQRMKPGVYTRLPFLFDWKENT			681
Db	387 SGGPLVCEPSGSRFPLAGIVSWGICGAERRRGVAVTRLDWILIENT			435

Db	154	CPGNSFCGNSQCVTKVNPBECDDQDCSDGSDSEAHCEGLQAPAMRWAGRIVGMEAS	PGE	213
Qy	454	MPWQVSLHALGCGHICGASLISFPMLVLSAAHCYIDDRGRYSDPTQWTFGLGHDSQS	513	
Db	214	FPWQVSLALE-MKEHFCGAAIINARVLSAAHCFNE-----FDDPTGVAIVYGATYLSGE	267	
Qy	514	APGVERRLRKRISSHPFNDFPFDYDIALLTLEKPEAEGSWMVRPICLPASHVFPAGKAI	573	
Db	268	ASTVA-AQGVQVYKHPILNRADTDADPDVAVELTSLPFRRIHQPVCLPAIHIFFPSKCC	326	
Qy	574	WYTWGHTQYGG-TGALILQKGEIRVINQTTCENTLLPQOITPRMNCVGLSGVDSCQD	632	
Db	327	LISGMYLKEDPLVKEPEVLQKATVEBLDQALCASLYGHSILTRMVCAGYLDGKVDSCQD	386	
Qy	633	SGGFLSSVEADGRITFQAGVSWKSGDCAGQNRKGVYTRLPFLFDWIKENT	681	
Db	387	SGGPLVCEPSPGSRFPLAGIVSWGICGAEARRPGVAVRYTRLDWILEAT	435	
RESULT 23				
ENTK_BOVIN	ID	ENTK_BOVIN	STANDARD;	PRT; 1035 AA.
AC	P98072;			
DT	01-FEB-1996	(Rel. 33, Created)		
DT	01-FEB-1996	(Rel. 33, last sequence update)		
DT	01-OCT-2004	(Rel. 45, last annotation update)		
DE	Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).			
GN	Name=PRSS7; Synonyms=ENTK;			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	TISSUE=Ductodenum;			
RX	MEDLINE=94329561; PubMed=8052624;			
RA	Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;			
RT	"Enterokinase, the initiator of intestinal digestion, is a mosaic			
RT	protease composed of a distinctive assortment of domains.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).			
RN	[2]			
RP	SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=94043122; PubMed=8226855;			
RA	Lavallie E.R., Rehentulla A., Racie L.A., DiBlasio E.A., Ferez C.,			
RA	Grant K.L., Light A., McCoy J.M.;			
RT	"Cloning and functional expression of a cDNA encoding the catalytic			
RT	subunit of bovine enterokinase.";			
RL	J. Biol. Chem. 268:23311-23317(1993).			
RN	[3]			
RP	SEQUENCE OF 801-827.			
RC	TISSUE=Intestine;			
RX	MEDLINE=92189715; PubMed=1799406;			
RA	Light A., Janaka H.;			
RT	"The amino-terminal sequence of the catalytic subunit of bovine			
RT	enterokinase.";			
RL	J. Protein Chem. 10:475-480(1991).			
CC	-1- FUNCTION: Responsible for initiating activation of pancreatic			
CC	proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase			
CC	A). It catalyzes the conversion of trypsinogen to trypsin which in			
CC	turn activates other proenzymes including chymotrypsinogen,			
CC	carboxypeptidases, and proelastases.			
CC	-1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys- -Ile-7 bond in			
CC	trypsinogen.			
CC	-1- SUBUNIT: Heterodimer of a catalytic (light) chain and a			
CC	multidomain (heavy) chain linked by a disulfide bond.			
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).			
CC	-1- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=2;			
CC	Name=Long;			
CC	IsoId=P98072-1; Sequence=Displayed;			
CC	Name=Short;			
CC	IsoId=P98072-2; Sequence=VSD_005386;			

CC -1- TISSUE SPECIFICITY: Intestinal brush border.  
 CC -1- PM: The chains are derived from a single precursor that is  
 CC cleaved by a trypsin-like protease.  
 CC -1- SIMILARITY: Belongs to peptidase family S1.  
 CC -1- SIMILARITY: Contains 2 CUB domains.  
 CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.  
 CC -1- SIMILARITY: Contains 1 SPA domain.  
 CC -1- SIMILARITY: Contains 1 SRCR domain.  
 CC -1- SIMILARITY: Contains 1 SRCR domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial  
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; U09859; AAB40026.1; -;  
 DR EMBL; L19663; AAB40035.1; -;  
 DR PIR; A43090; A43090.  
 DR MEROPS; S01.156; -;  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR002172; LDL\_receptor\_A.  
 DR InterPro; IPR000998; MAM.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR InterPro; IPR011163; Pept\_S1A\_enterop.  
 DR InterPro; IPR009003; Pept\_Ser\_Cys.  
 DR InterPro; IPR000082; SRA.  
 DR InterPro; IPR001190; Srcr\_receptor.  
 DR Pfam; PF00431; CUB; 2.  
 DR Pfam; PF00057; Ldl\_recept\_a; 2.  
 DR Pfam; PF00629; MAM; 1.  
 DR Pfam; PF01390; SEA; 1.  
 DR Pfam; PF00510; SRCR; 1.  
 DR Pfam; PF00083; Trypsin; 1.  
 DR PIRSF; PIRSF001138; Enteropeptidase; 1.  
 DR PRINTS; PRO0722; CHYMOTRYPSIN.  
 DR PRINTS; PRO0261; LDLRECEPTOR.  
 DR PROSITE; PS01180; CUB; 2.  
 DR PROSITE; PS01209; LDLRA\_1; 2.  
 DR PROSITE; PS50068; LDLRA\_2; 2.  
 DR PROSITE; PS00740; MAM\_1; 1.  
 DR PROSITE; PS50060; MAM\_2; 1.  
 DR PROSITE; PS50024; SEA; 1.  
 DR PROSITE; PS00420; SRCR\_2; 1.  
 DR PROSITE; PS50287; SRCR\_1; FALSE\_NEG.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM 3D-structure; Alternative splicing; Direct protein sequencing;  
 KM Glycoprotein; Hydrolase; Lipoprotein; Myristate; Repeat;  
 KM Serine protease; Signal-anchor; Transmembrane; Zymogen.  
 FT CHAIN 1 800 Enteropeptidase non-catalytic chain.  
 FT CHAIN 1 1035 Enteropeptidase catalytic chain.  
 FT DOMAIN 1 18 Cytoplasmic (Potential).  
 FT TRANSMEM 19 47 Signal-anchor for type II membrane  
 FT protein (Potential).  
 FT DOMAIN 48 1035 Extracellular (Potential).  
 FT DOMAIN 54 169 SEA.  
 FT DOMAIN 197 238 LDL-receptor class A 1.  
 FT DOMAIN 240 350 CUB 1.  
 FT DOMAIN 358 520 MAM.  
 FT DOMAIN 540 650 CUB 2.  
 FT DOMAIN 657 695 LDL-receptor class A 2.  
 FT DOMAIN 694 787 SRCR.  
 FT DOMAIN 801 1035 Serine protease.  
 FT ACT\_SITE 841 841 Charge relay system (By similarity).  
 FT ACT\_SITE 892 892 Charge relay system (By similarity).  
 FT ACT\_SITE 987 987 Charge relay system (By similarity).  
 FT LITID 2 2 N-myristoyl glycine (Potential).

FT DISULFID 199 212 By similarity.  
 FT DISULFID 206 225 By similarity.  
 FT DISULFID 219 236 By similarity.  
 FT DISULFID 659 671 By similarity.  
 FT DISULFID 666 684 By similarity.  
 FT DISULFID 678 693 By similarity.  
 FT DISULFID 788 812 Interchain (By similarity).  
 FT DISULFID 826 842 By similarity.  
 FT DISULFID 926 933 By similarity.  
 FT DISULFID 937 972 By similarity.  
 FT DISULFID 983 1011 By similarity.  
 FT CARBOHYD 116 116 N-linked (GlcNAc...)  
 FT CARBOHYD 147 147 N-linked (GlcNAc...)  
 FT CARBOHYD 170 170 N-linked (GlcNAc...)  
 FT CARBOHYD 194 194 N-linked (GlcNAc...)  
 FT CARBOHYD 233 233 N-linked (GlcNAc...)  
 FT CARBOHYD 263 263 N-linked (GlcNAc...)  
 FT CARBOHYD 264 264 N-linked (GlcNAc...)  
 FT CARBOHYD 404 404 N-linked (GlcNAc...)  
 FT CARBOHYD 456 456 N-linked (GlcNAc...)  
 FT CARBOHYD 486 486 N-linked (GlcNAc...)  
 FT CARBOHYD 519 519 N-linked (GlcNAc...)  
 FT CARBOHYD 550 550 N-linked (GlcNAc...)  
 FT CARBOHYD 646 646 N-linked (GlcNAc...)  
 FT CARBOHYD 698 698 N-linked (GlcNAc...)  
 FT CARBOHYD 722 722 N-linked (GlcNAc...)  
 FT CARBOHYD 741 741 N-linked (GlcNAc...)  
 FT CARBOHYD 762 762 N-linked (GlcNAc...)  
 FT CARBOHYD 864 864 N-linked (GlcNAc...)  
 FT CARBOHYD 903 903 N-linked (GlcNAc...)  
 FT CARBOHYD 965 965 N-linked (GlcNAc...)  
 FT VARSPIC 166 192 Missing (in isoform Short).  
 FT CONFLICT 808 808 /FTID=VSP\_005386.  
 FT STRAND 802 802 R -> Y (in Ref. 3).  
 FT TURN 805 806  
 FT TURN 809 810  
 FT TURN 813 814  
 FT STRAND 815 820  
 FT TURN 821 822  
 FT STRAND 823 830  
 FT STRAND 835 838  
 FT HELIX 840 843  
 FT TURN 844 845  
 FT HELIX 850 852  
 FT STRAND 853 857  
 FT TURN 861 861  
 FT STRAND 862 863  
 FT TURN 868 869  
 FT STRAND 871 880  
 FT TURN 882 883  
 FT STRAND 885 885  
 FT TURN 886 889  
 FT STRAND 890 890  
 FT STRAND 894 898  
 FT STRAND 912 912  
 FT TURN 916 917  
 FT TURN 922 923  
 FT STRAND 925 930  
 FT STRAND 933 933  
 FT TURN 936 937  
 FT STRAND 940 940  
 FT STRAND 943 943  
 FT STRAND 945 951  
 FT HELIX 954 960  
 FT TURN 962 963  
 FT TURN 968 969  
 FT TURN 970 973  
 FT TURN 976 977  
 FT STRAND 981 981  
 FT STRAND 984 985  
 FT TURN 987 988  
 FT STRAND 990 995



```

FT DOMAIN 579 614 LDL-receptor class A 5.
FT DOMAIN 615 653 LDL-receptor class A 6.
FT DOMAIN 654 690 LDL-receptor class A 7.
FT DOMAIN 690 786 SRCR.
FT ACT_SITE 802 1042 Serine protease.
FT ACT_SITE 843 943 Charge relay system (By similarity).
FT ACT_SITE 892 892 Charge relay system (By similarity).
FT ACT_SITE 985 985 Charge relay system.
FT DISULFID 790 912 By similarity.
FT DISULFID 828 844 By similarity.
FT DISULFID 955 970 By similarity.
FT DISULFID 981 1010 By similarity.
FT CARBOHYD 80 80 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 104 104 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 135 135 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 231 231 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 245 245 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 251 251 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 305 305 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 320 320 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 376 376 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 413 413 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 446 446 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 451 451 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 469 469 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 567 567 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 651 651 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 697 697 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 761 761 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1022 1022 N-linked (GlcNAc...) (Potential).
FT MUTAGEN 985 985 S->A: Loss of activity.
FT CONFLICT 854 854 W->R (in Ref. 2).
FT CONFLICT 876 876 K->R (in Ref. 2).
SQ SEQUENCE 1042 AA; 116564 MW; 7705398BB607AD2 CRC64;

```

Query March 18.3%; Score 694; DB 1; Length 1042;  
 Best Local Similarity 28.6%; Pred. No. 3.3e-40;  
 Matches 214; Conservative 95; Mismatches 255; Indels 180; Gaps 34;

```

QY 30 DSKTVORTDNCSGFGHARGVELMFTTGPFPDPSYPYAHARCOMALRGDAD----- 81
DB 365 DHDCVDSDEVNCS--CHSGGLVECR-----NGCIPSTFQCD--GDDCKDGDGDEE 412
QY 82 --SYLSLTFNSFD-----LACDERGSLVTVVNTL---SPMEPALVQLGTYPPS 128
DB 413 NCSYIOTSQCGDRCILYNPCLDSCG--GSSILCDPNNSLNCSOCEPITL-ELCMNLP-- 467
QY 129 YNLTFHSSQNVLLITLITNTERRRHGFATFQLPRMSSCGRLRK-AQGTFNSPYVPGH 187
DB 468 YNST-----STPNY-----FGRRTQKESISWESSLPFAL 497
QY 188 YRPN-----IDCTWNIENVNNOHVKKVRFKFLLEPRRACGTCPKDYVE-----INGEK 236
DB 498 VQTNCKYKIMFPSC--ILVE-----KCDVNTGERIRIPCRALCHSKKRCGSVIGIVLQ 550
QY 237 YCGER--SQFVYVNSNKITVRFHSDOSYTDGTGLAETLSYDSDPC-PGQFTCTGRCI 293
DB 551 WPEPDTDSQFPEENSNDQTCI-----WPEDEV-----EECSPSHFKCRSGQCV 593
QY 294 RKELRCSGMACCTHSDSLNCSDAHQFTC-KKKFKCFLFWVDSVVDGDNDEQGS 352
DB 594 LASRRCDGQADCDSDSEENCGCKERRDLMECPNNQCLAKTIVICDGFPCDPYDEKNGS 653
QY 353 -CPAQTFRCNSGKCLSKSQCCNGKDDCCGSDDEASCPKVVN----- 392
DB 654 FCGDDELECAHACVSRDLWCDEGLDSDSDMDCVTLSTINWSSSFLWVHRAATEHHV 713
QY 393 -----VTC-----TKAT-YRCLNGCLSK--GNP 413
DB 714 CADGMOBITSLQACKQWGLGEPVTKLIOEQEKPRWTLTHSNWESLGTTLHRLVNGQ 773
QY 414 ECDKEDCSDSDSDKDCDCCGGRSFTF-QARVVGCTDADDEGMPQVSLHALGQGHICGS 472

```

```

DB 774 SCESRSKISLTCMQ--DCGRPARARMKRIILGRTSRPGRWPMQCSLQSPSGHICGV 831
QY 473 LISPNWLVSAAHCTYDGRGRYSPDTQWTAFLGLHDOSQSAPOVQERRLKRIISHPPFN 532
DB 832 LIACKWTLVAHCP-EER---ENAAVWKVVLGINNDHPSV-FMQRFPVYTTILHPRYS 885
QY 533 DFTFDYIALLELEKPAVSVMRPICLPDASHVFPAGKAIWYMGHTQYGGTALILQ 592
DB 886 RAVVDYISIVELSEDSISFTGYVAPVCLPMBQMLEPDTYCYITGMCH--MGKMPFKIQ 943
QY 593 KGEIRVINTTCENILPQO--ITPRMVCVFLSGGVDSGCGDGGGLSSVADGRIFQAGV 651
DB 944 EGEVRILSLHRCQSPFMKITTMTICAGYESGTVDSGMDSGGSLVCEKRGKWTLLFGL 1003
QY 652 VSWGDGAQR-NKRGVYTRPLPRDWIK 678
DB 1004 TSMGSCVCFSKVLGPGVSNVSYFVEMWK 1031

```

RESULT 25  
 ID Q80YN4 PRELIMINARY; PRT; 1111 AA.  
 AC Q80YN4;  
 DT 01-JUN-2003 (Tremblrel, 24, Created)  
 DT 01-JUN-2003 (Tremblrel, 24, Last sequence update)  
 DT 01-MAR-2004 (Tremblrel, 26, Last annotation update)  
 DE Corin.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,  
 OC Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA Langenickel T., Pagel I., Buttgeriet J., Tenner K., Lindner M.,  
 RL Willenbrock R., Dietz R., Bader M.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 CC 1- SIMILARITY: Belongs to peptidase family S1.  
 DR EMBL; AY251285; A086772.1; -.  
 DR HSSP; Q07954; 1CR8.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0007275; P:development; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000024; Fz domain.  
 DR InterPro; IPR002172; LDL receptor A.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR InterPro; IPR009003; Pept\_Ser\_Cys.  
 DR Pfam; PF01392; Fz; 2.  
 DR Pfam; PF00057; Ldl\_recept\_a; 6.  
 DR Pfam; PF00089; Trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPIN.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR SMART; SM00063; FRI; 2.  
 DR SMART; SM00192; IDLA; 7.  
 DR SMART; SM00202; SR; 1.  
 DR SMART; SM00202; TRYP\_SRC; 1.  
 DR PROSITE; PS0038; Fz; 2.  
 DR PROSITE; PS01209; LDLRA\_1; 6.  
 DR PROSITE; PS0068; LDLRA\_2; 7.  
 DR PROSITE; PS50287; SRCR\_2; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 1111 AA; 122635 MW; 3BA2706CBB81157F CRC64;  
 Query March 18.2%; Score 690.5; DB 2; Length 1111;  
 Best Local Similarity 28.4%; Pred. No. 6.2e-40;

Matches 210; Conservative 79; Mismatches 252; Indels 199; Gaps 32;

QY 33 TQVGTQNGSGFGLHARGVELMRTTGGFPDPSYPFARACQMALRGADSVLSTFRSPD 92  
 DB 462 TFGDGEDCKDGDSENCSS-----DRPRP-----CPGGRGLDS----- 497

QY 93 LASGDE--RGSADVTVNTL---SPMEPHALVQLGCTPYPSYNTLTFP----- 134  
 DB 498 --SVESECASSSLCDSSSLNSCHCEPITL-ELQNLPL--YNLTHPNVLGHRTOKEAS 552

QY 135 -SSQNVLLITLITLINTERRHPEFATFPQLPRMSSCGRLRAQGTENSPYPPHYPP-NI 192  
 DB 553 ISMESALFPALVQNTCYKYLMEFFACTLIVPKCDVNTQ-----RVPPCRL 597

QY 193 DCTWNIEVPPNNOHKYAFKFPYLLPERRAGTC-----PKDYVEINERKVGGEESQF 244  
 DB 598 LC-----EHSKER-----CESVIGVGLQMPED-----TDSQF 626

QY 245 VVTSNSNKITVRFHSDQSYTDGFLAAYLSYSSDPC-PCQFTGTCRCIRKELRCDMA 303  
 DB 627 PEGSSDNQTCL-----LPNEDEVESPSHFKRSGRCVLGSRCGQGA 669

QY 304 DCTDHSDELNCSADAGHQFTCK-NKFKCPFLWCDVNDGDSNDEQCS-CPAQTRCS 361  
 DB 670 DCDSDSDEENCGCKERDLMECPILKQCLKHTLIDGFPDSDSMDERNKCSFQDDELECA 729

QY 362 NGKLSAQCGNGKDDCGDSDSEASCPKAVNYTCTKH-----TYRCING- 405  
 DB 730 NHECVPRDLNCDGWTDCSDSDSEWGC-----VTLSSKNGSSSPLTHRSARDHHVCADGW 784

QY 406 -----LC--LSKGNPE-----CDGKE-----DCSDGS 425  
 DB 785 QETLSQACQMGEGPESVTELVGQGGQQLRHSSWENLNGSTLQELVHRSCPSGS 844

QY 426 DEK---DCDGLRSFTR-QARVVGCTDADGEMPMQVSLHALCGHICGASLISPMVLV 480  
 DB 845 EISLCTKQCCGRPARMRKRIIGGRTSRPGRMWOCOSLSEPSGHCIGCVLITAKKWL 904

QY 481 SAACHYIDDDGFRSDPTQWTAFLGLHDQGRSAPGVQERLKRITISHPFNDPTFPYDI 540  
 DB 905 TVAAHCF-EGR---EDADVNVVPGINNLDHPSC-FMQTRVKTLLHPFRSRAVVYDI 958

QY 541 ALLELEPAEYSSWVRPCLPDASHVPAKAIWVTGHTQYGGTALILQKEIRVIN 600  
 DB 959 SVVSLSDINDIMETSVYRVCLEPREFLEPDYCYITWGH--WKNKMPFLQBEVAILIP 1016

QY 601 QTTCENLIPQO-ITPRMNCVGLSGVDSQCGSGGSPUSVEADGRIPLQGVSWSDGCA 659  
 DB 1017 LEOCOSYFDMKTIITNMICAGYESGTVDSQWSDGSPVCEBPGQWTLFGLTWSGVCF 1076

QY 660 QR-NKPGVYTRLPFLRDMIK 678  
 DB 1077 SKVLGPGVYSNVSYFVDMIE 1096

RESULT 26

08CANG PRELIMINARY; PRT; 777 AA.

AC 08CANG;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Mus musculus 16 days neonate thymus cDNA, RIKEN full-length enriched library, clone:Al30097D21 product:protease, serine, 7 (enterokinase), full insert sequence. (fragment).  
 DE Name=Prs87;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;

RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning."; Meth. Enzymol. 303:19-44(1999).  
 RL [2]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection."; Nature 409:685-690(2001).  
 RL [3]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
 RA The FANTOM Consortium.  
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."; Nature 420:563-573(2002).  
 RL [4]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes."; Genome Res. 10:1617-1630(2000).  
 RL [5]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Konno H., Akiyama U., Nishi K., Katsunai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kasaiwaagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuyama S., Kawai U., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multiplexed capillary sequencer."; Genome Res. 10:1757-1771(2000).  
 RL [6]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
 RA Aichi U., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P., Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T., Hori F., Imclanti K., Ishii Y., Itoh M., Kagawa I., Kasukawa T., Katoh H., Kawai U., Kojima Y., Kondo S., Konno H., Kouda M., Koya S., Kuribara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M., Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y., Saito R., Satoh H., Sakai C., Sakai K., Sakazume N., Sano H., Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M., Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T., Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RL  
 CC -I- SIMILARITY: Belongs to peptidase family S1.  
 CC -I- SIMILARITY: Contains 1 MAM domain.  
 DR EMBL; AK038356; BAC29973.1; -.  
 DR MGD; MGI:1197523; Prs87.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0016301; F:kinase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR002172; IDL\_receptor\_A.  
 DR InterPro; IPR000998; MAM.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.

DR InterPro: IPR009003; Pept. Ser. Cys.  
 DR InterPro: IPR001190; SerCr\_receptor.  
 DR Pfam: PF00431; CUB; 2.  
 DR Pfam: PF00057; Ldl\_recept\_a; 1.  
 DR Pfam: PF00629; MAM; 1.  
 DR Pfam: PF00530; SRCR; 1.  
 DR Pfam: PF00089; Trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PRINTS: PR00020; MAMOMAIN.  
 DR SMART: SM00042; CUB; 1.  
 DR SMART: SM00192; LDLa; 1.  
 DR SMART: SM00137; MAM; 1.  
 DR SMART: SM00202; SR; 1.  
 DR SMART: SM00020; Tryp\_Spc; 1.  
 DR PROSITE: PS01180; CUB; 1.  
 DR PROSITE: PS01209; LDLa\_1; 1.  
 DR PROSITE: PS00068; LDLa\_2; 1.  
 DR PROSITE: PS00740; MAM\_1; UNKNOWN\_1.  
 DR PROSITE: PS00600; MAM\_2; 1.  
 DR PROSITE: PS0287; SRCR\_2; 1.  
 DR PROSITE: PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 DR GlycoProtein: Hydrolase; Kinase; Protease; Serine protease.  
 DR NON\_TER 1  
 SC SEQUENCE 777 AA; 87314 MW; A18E2F4ECF06D3A8 CRC64;  
 Query Match 18.2%; Score 688; DB 2; Length 777;  
 Best Local Similarity 29.3%; Pred. No. 6.2e-40;  
 Matches 186; Conservative 93; Mismatches 220; Indels 136; Gaps 22;  
 QY 86 LTFSSFDLASCDEGSDLVYVYNTLSPEPHALVOLGCTYPPSNLTHTSSQNVLLTLI 145  
 DB 231 VVFAAFRRNGSTALDIDSLTNGICSGSPYPEPTLVTPPP----- 272  
 QY 146 TINTERRRHPFEATFQFOLPRMSSCGG--RLRAQGTFFNSPYRGHYPPMIDCTMNIEVNN 203  
 DB 273 -----ELP--TDCCGFPELMFNSTFSSFPNDKYPNASCITMILNMRG 315  
 QY 204 QHVAVRKFVYLLPRRACGTCPPDYVEI--NGE-----KYGERSQFVNTSNKIT 254  
 DB 316 KNIGLHGFLENDL-----INDVYVRDGEFDSLALLVYTGPGVKDLFTYTRMT 367  
 QY 255 VRPHSDQSYDTGTGLAEYLS---YDSDPC--PGQFTCTGTCIRKEKLCDCMACTDHS 310  
 DB 368 VIFLTNNETRRKRGKAFNTSGYVIGIPPCDDDFOCDCNGCIPLGMLCDSPYPCRDSD 427  
 QY 311 ELNC-----SCDAGHQFTCKAKKFKPLFMVCDVNDGDNDSBQSCSPAOTFRCSN 362  
 DB 428 EASCVRFLNGTRNSNGLVQFINHS-----IWHI-----ACAEV-----WTTQISN 467  
 QY 363 GKCLSKSQCCNGKDCDGDSDDEASC-----KVVNVTCYKHYRCINGLCLSKGN 412  
 DB 468 EVC-----HLLGLGSANSSWPISSSTGGPFVRN-----QAVNGSLILTPS 508  
 QY 413 PECCKE---DCSDGSDKDCDGLASFTROA--RVGCTDABEGEFPVOYSLHALGCG 466  
 DB 509 LQCSQDSLILLCQCHKMS-----CGEKVTKQVSPKTYGGSDAQAQAMPVVALYHRRS 562  
 QY 467 H---ICGASLISPMVLVAACHYIDDRFRSDPQWAFGLHDSQSRASPAQVERLTK 523  
 DB 563 TDRLICGASLVSQDLVLSAACHYVR---RNLDPFRKRAVLGLHQSULTSPOVRRVVD 618  
 QY 524 RIISHPPENDTFYDIALLELEKPAEYSSMWRLPCLDASHVPAGKAIWVGWGHATQY 583  
 DB 619 QIVINPHYDRARKVNDIMWHLFKNVYTDVITQICPEBENQIFIPGRTCSIAAGGYDKI 678  
 QY 584 GGTGALLIQKEIRIVYNOTTCENLIPQ-QITPRMKCVFLSGVDSCCGSDGSPISSTVEA 642  
 DB 679 NGSTVDVLEKADVPLISNEKCOQOLPEYNLTESMTCAGYEBGGIDSCGSDGSPLMCOE- 737  
 QY 643 DGRIFQAGVSMGDCQACRANKRGVYTRPLFRDWI 677

DB 738 NNRWFLVGTSTFGVQCALPNHGVYRVVSQFIEMI 772  
 RESULT 27  
 ID ENTX HUMAN STANDARD; PRT; 1019 AA.  
 AC P96073;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Enteroleptidase precursor (EC 3.4.21.9) (Enterokinase).  
 GN Name=PRSS7; Synonyms=ENTX;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Duodenum;  
 RX MEDLINE=95234679; PubMed=7718557;  
 RA Kitamoto Y., Veille R.A., Donis-Keller H., Sadler J.E.;  
 RT "cDNA sequence and chromosomal localization of human enterokinase, the  
 RT proteolytic activator of trypsinogen."  
 RL Biochemistry 34:4562-4568(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND DISEASE.  
 RX MEDLINE=21606074; PubMed=11719902;  
 RA Holzinger A., Maier E.M., Buck C., Mayerhofer P.U., Kappler M.,  
 RA Haworth J.C., Moroz S.P., Hadorn H.-B., Sadler J.E., Roscher A.A.;  
 RT "Mutations in the proenteropeptidase gene are the molecular cause of  
 RT congenital enteropeptidase deficiency."  
 RL Am. J. Hum. Genet. 70:20-25(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20289799; PubMed=10830953;  
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
 RA Park H.-S., Toyoda A., Ishii K., Tokoi Y., Choi D.-K., Groner Y.,  
 RA Soeda E., Ohtsuki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,  
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Paterson D.,  
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
 RA Rosenblatt A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
 RA Shienrai A., Sasaki T., Nagamine K., Mitsuyama S., Antocaklis S.E.,  
 RA Moshima S., Shimizu N., Nordstiek G., Hornisiek G., Brandt P.,  
 RA Scharte M., Schen O., Desario A., Reichelt J., Kauter G., Blocker H.,  
 RA Ramser U., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,  
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
 RA Lehrach H., Reinhardt R., Yaepp M.-D.;  
 RT "The DNA sequence of human chromosome 21."  
 RL Nature 405:311-319(2000).  
 RN [4]  
 RP SEQUENCE OF 749-1019 FROM N.A.  
 RC TISSUE=Duodenum;  
 RX MEDLINE=94329561; PubMed=8052624;  
 RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;  
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic  
 RT protease composed of a distinctive assortment of domains."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).  
 CC -1- FUNCTION: Responsible for initiating activation of pancreatic  
 CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase  
 CC A). It catalyzes the conversion of trypsinogen to trypsin which in  
 CC turn activates other proenzymes including chymotrypsinogen,  
 CC procarboxypeptidases, and proelastases.  
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in  
 CC trypsinogen.  
 CC -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a  
 CC multidomain (heavy) chain linked by a disulfide bond.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -1- TISSUE SPECIFICITY: Intestinal brush border.  
 CC -1- PTM: The chains are derived from a single precursor that is  
 CC cleaved by a trypsin-like protease.  
 CC -1- DISEASE: Defects in PRSS7 are a cause of enterokinase deficiency  
 CC (MIM:226200); a life-threatening intestinal malabsorption disorder  
 CC characterized by diarrhea and failure to thrive.



```

CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 MAM domain.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, U09860, AAC50138.1, -.
DR EMBL, Y19124, CAB65555.1, -.
DR EMBL, Y19125, CAB65555.1, JOINED.
DR EMBL, Y19126, CAB65555.1, JOINED.
DR EMBL, Y19127, CAB65555.1, JOINED.
DR EMBL, Y19128, CAB65555.1, JOINED.
DR EMBL, Y19129, CAB65555.1, JOINED.
DR EMBL, Y19130, CAB65555.1, JOINED.
DR EMBL, Y19131, CAB65555.1, JOINED.
DR EMBL, Y19132, CAB65555.1, JOINED.
DR EMBL, Y19133, CAB65555.1, JOINED.
DR EMBL, Y19134, CAB65555.1, JOINED.
DR EMBL, Y19135, CAB65555.1, JOINED.
DR EMBL, Y19136, CAB65555.1, JOINED.
DR EMBL, Y19137, CAB65555.1, JOINED.
DR EMBL, Y19138, CAB65555.1, JOINED.
DR EMBL, Y19139, CAB65555.1, JOINED.
DR EMBL, Y19140, CAB65555.1, JOINED.
DR EMBL, Y19141, CAB65555.1, JOINED.
DR EMBL, Y19142, CAB65555.1, JOINED.
DR EMBL, Y19143, CAB65555.1, JOINED.
DR EMBL, AL163218, CAB90392.1, -.
DR EMBL, AL163217, CAB90389.1, -.
DR PIR, A56318, A56318.
DR HSSP, P98072, 1EKB.
DR MEROPS, S01_156, -.
DR GENE, HGNC:9490, PRSS7.
DR MIM, 606635, -.
DR GO, GO:0005903, C:brush border, TAS.
DR InterPro, IPR000859, CUB.
DR InterPro, IPR002172, LDL_receptor_A.
DR InterPro, IPR000998, MAM.
DR InterPro, IPR001354, Peptidase_S1.
DR InterPro, IPR001114, Peptidase_S1A.
DR InterPro, IPR001163, Pept_S1A_enterop.
DR InterPro, IPR009003, Pept_Ser_Cys.
DR InterPro, IPR000082, SEA.
DR Pfam, PF000431, CUB, 2.
DR Pfam, PF00057, LDL_recept_a, 2.
DR Pfam, PF00629, MAM, 1.
DR Pfam, PF01390, SEA, 1.
DR Pfam, PF00530, SRCC, 1.
DR Pfam, PF00089, Trypsin, 1.
DR Pfam, PF000118, Enteropeptidase, 1.
DR PRINTS, PR00722, CHYMOTRYPSIN.
DR PRINTS, PR00261, LDLRECEPTOR.
DR PRINTS, PR00020, MAMDOMAIN.
DR PROSITE, PS01180, CUB, 2.
DR PROSITE, PS01209, LDLRA_1, 2.
DR PROSITE, PS00068, LDLRA_2, 2.
DR PROSITE, PS00740, MAM_1, 1.
DR PROSITE, PS00060, MAM_2, 1.
DR PROSITE, PS00024, SEA, 1.
DR PROSITE, PS00420, SRCC_1, FALSE_NEG.
DR PROSITE, PS00287, SRCC_2, 1.
DR PROSITE, PS50240, TRYPSIN_DOM, 1.

```

DR	PROSITE: PS00134; TRYPSIN_HIS: 1.	DR	PROSITE: PS00135; TRYPSIN_SER: 1.	DR	PROSITE: PS00136; TRYPSIN_SER: 1.
KM	Glycoprotein; Hydrolyase; Lipoprotein; Myristate; Repeat;	KM	Glycoprotein; Hydrolyase; Lipoprotein; Myristate; Repeat;	KM	Glycoprotein; Hydrolyase; Lipoprotein; Myristate; Repeat;
KM	Serine protease; Signal-anchor; Transmembrane; Zymogen.	KM	Serine protease; Signal-anchor; Transmembrane; Zymogen.	KM	Serine protease; Signal-anchor; Transmembrane; Zymogen.
FT	CHAIN 1 784	FT	CHAIN 1 784	FT	CHAIN 1 784
FT	CHAIN 1 784	FT	CHAIN 1 784	FT	CHAIN 1 784
FT	CHAIN 1 784	FT	CHAIN 1 784	FT	CHAIN 1 784
FT	TRANSMEM 19 47	FT	TRANSMEM 19 47	FT	TRANSMEM 19 47
FT	DOMAIN 48 1019	FT	DOMAIN 48 1019	FT	DOMAIN 48 1019
FT	DOMAIN 52 169	FT	DOMAIN 52 169	FT	DOMAIN 52 169
FT	DOMAIN 182 223	FT	DOMAIN 182 223	FT	DOMAIN 182 223
FT	DOMAIN 225 334	FT	DOMAIN 225 334	FT	DOMAIN 225 334
FT	DOMAIN 342 504	FT	DOMAIN 342 504	FT	DOMAIN 342 504
FT	DOMAIN 524 634	FT	DOMAIN 524 634	FT	DOMAIN 524 634
FT	DOMAIN 641 679	FT	DOMAIN 641 679	FT	DOMAIN 641 679
FT	DOMAIN 678 771	FT	DOMAIN 678 771	FT	DOMAIN 678 771
FT	DOMAIN 785 1019	FT	DOMAIN 785 1019	FT	DOMAIN 785 1019
FT	ACT SITE 825 825	FT	ACT SITE 825 825	FT	ACT SITE 825 825
FT	ACT SITE 876 876	FT	ACT SITE 876 876	FT	ACT SITE 876 876
FT	ACT SITE 971 971	FT	ACT SITE 971 971	FT	ACT SITE 971 971
FT	LIPID 2 2	FT	LIPID 2 2	FT	LIPID 2 2
FT	DISULFID 184 197	FT	DISULFID 184 197	FT	DISULFID 184 197
FT	DISULFID 191 210	FT	DISULFID 191 210	FT	DISULFID 191 210
FT	DISULFID 204 221	FT	DISULFID 204 221	FT	DISULFID 204 221
FT	DISULFID 643 655	FT	DISULFID 643 655	FT	DISULFID 643 655
FT	DISULFID 668 668	FT	DISULFID 668 668	FT	DISULFID 668 668
FT	DISULFID 662 677	FT	DISULFID 662 677	FT	DISULFID 662 677
FT	DISULFID 772 826	FT	DISULFID 772 826	FT	DISULFID 772 826
FT	DISULFID 810 826	FT	DISULFID 810 826	FT	DISULFID 810 826
FT	DISULFID 910 977	FT	DISULFID 910 977	FT	DISULFID 910 977
FT	DISULFID 941 956	FT	DISULFID 941 956	FT	DISULFID 941 956
FT	CARBOHYD 116 116	FT	CARBOHYD 116 116	FT	CARBOHYD 116 116
FT	CARBOHYD 147 147	FT	CARBOHYD 147 147	FT	CARBOHYD 147 147
FT	CARBOHYD 179 179	FT	CARBOHYD 179 179	FT	CARBOHYD 179 179
FT	CARBOHYD 328 328	FT	CARBOHYD 328 328	FT	CARBOHYD 328 328
FT	CARBOHYD 335 335	FT	CARBOHYD 335 335	FT	CARBOHYD 335 335
FT	CARBOHYD 388 388	FT	CARBOHYD 388 388	FT	CARBOHYD 388 388
FT	CARBOHYD 440 440	FT	CARBOHYD 440 440	FT	CARBOHYD 440 440
FT	CARBOHYD 470 470	FT	CARBOHYD 470 470	FT	CARBOHYD 470 470
FT	CARBOHYD 503 503	FT	CARBOHYD 503 503	FT	CARBOHYD 503 503
FT	CARBOHYD 534 534	FT	CARBOHYD 534 534	FT	CARBOHYD 534 534
FT	CARBOHYD 630 630	FT	CARBOHYD 630 630	FT	CARBOHYD 630 630
FT	CARBOHYD 682 682	FT	CARBOHYD 682 682	FT	CARBOHYD 682 682
FT	CARBOHYD 706 706	FT	CARBOHYD 706 706	FT	CARBOHYD 706 706
FT	CARBOHYD 725 725	FT	CARBOHYD 725 725	FT	CARBOHYD 725 725
FT	CARBOHYD 848 848	FT	CARBOHYD 848 848	FT	CARBOHYD 848 848
FT	CARBOHYD 887 887	FT	CARBOHYD 887 887	FT	CARBOHYD 887 887
FT	CARBOHYD 909 909	FT	CARBOHYD 909 909	FT	CARBOHYD 909 909
FT	CARBOHYD 949 949	FT	CARBOHYD 949 949	FT	CARBOHYD 949 949
FT	CONFLICT 134 134	FT	CONFLICT 134 134	FT	CONFLICT 134 134
FT	CONFLICT 732 732	FT	CONFLICT 732 732	FT	CONFLICT 732 732
FT	CONFLICT 754 771	FT	CONFLICT 754 771	FT	CONFLICT 754 771
SEQ	SEQUENCE 1019 AA; 112923 MM; B6AA245F6D4A563 CRC64;	SEQ	SEQUENCE 1019 AA; 112923 MM; B6AA245F6D4A563 CRC64;	SEQ	SEQUENCE 1019 AA; 112923 MM; B6AA245F6D4A563 CRC64;
Query Match	18.0%; Score 683.5; DB 1; Length 1019;	Query Match	18.0%; Score 683.5; DB 1; Length 1019;	Query Match	18.0%; Score 683.5; DB 1; Length 1019;
Beet Local Similarity	32.5%; Pred. No. 1.7e-39;	Beet Local Similarity	32.5%; Pred. No. 1.7e-39;	Beet Local Similarity	32.5%; Pred. No. 1.7e-39;
Matches 180; Conservative 80; Mismatches 202; Indels 91; Gaps 20;		Matches 180;			

```

QY 123 TCNK----KFCCKPLFWCDVNDGDNDEQSCSPAOTFRCSNGKLSKQCGCKDCC 378
D 684 TTANNGLVFRIGIOW----HTACAE-----WTTOISNDVC-----QLL 719
QY 379 GDGSDASCPKRVNVTCTKHTYRCLANGCLSKGNECDG-----KEDCSDGS-----D 426
D 720 GLSGGNSKSPFST-----DGGPFVKNTAPDGHILNPSQCLDLSIRLCN 768
QY 427 EKDDCGLRSTFRARVVGSTADDEGEPPVYSLHALGQGHICGASLSPWLYSAHCV 486
D 769 HKSGGKTLAADIPTPKIVGSGNAKEGAMPVWGVY-YGRLLCASLVSDDMLVSAHCV 827
QY 487 IDDGFPYDPTQWTAFLGLDGSORSAPOGVERLKRRIISHPFNDPTFYDALTELE 546
D 828 YG-----RNLERSKATLILGLMKSNLTSPOVPLIDEIVNPHYRRKNDNLAMHLE 883
QY 547 KPAEYSMWRBICLPDASHVPAKAIWVTGHTOYGGTALILQKEIRVINQTCEN 606
D 884 FKVYTYDIQICLPENQVPPGRNCSIAGMGTIVVGSTANLLEADVPLSNRQOQ 943
QY 607 LLPQ-QITPRMVCVFGSGVSCQDGGSLSEVADGRIFQAGVSWGDCQQRNKP 665
D 944 CMPEYNTENMTCAGYERGGIDSCQDGGSLMCOE--NNRWFLAGVTSFGYKCALPVRPG 1002
QY 666 VYTRPLPFRMIX 678
D 1003 VYARVSRFTWIQ 1015

RESULT 28
ENTR_MOUSE STANDARD: PRT: 1069 AA.
ID ENTR_MOUSE
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Enteropепtidase (EC 3.4.21.9) (Enterokinasе).
GN Name=Prer87; Synonym=Entk;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN 1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Duoenum;
RA MEDLINE=98147142; PubMed=9486188;
RX Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;
RT "Structure of murine enterokinasе (enteropeptidase) and expression in
RT small intestine during development.";
RL Am. J. Physiol. 274:G342-G349(1998).
CC -1- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases (By similarity).
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
CC trypsinogen.
CC -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC multidomain (heavy) chain linked by a disulfide bond (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- PTM: The chains are derived from a single precursor that is
CC cleaved by a trypsin-like protease (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: U73378; AAB37317.1; -.
DR HSSP: P98072; 1EXB.
DR MEROPS: S01.156; -.
DR MED: MG1.1197523; Prer87.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR000998; MAM.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR InterPro: IPR001163; Pept_S1A_enterop.
DR InterPro: IPR009003; Pept_Ser_Cys.
DR InterPro: IPR001190; Srcr_receptor.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00057; LDL_recept_a; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF01390; SEA; 1.
DR Pfam: PF00530; SRCR; 1.
DR Pfam: PF00089; Trypsin; 1.
DR PIRSF: PIRSF001138; Enteropепtidase; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00192; LDLA; 2.
DR SMART: SM00137; MAM; 1.
DR SMART: SM00200; SEA; 1.
DR SMART: SM00282; SR; 1.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS00068; LDLRA_2; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS00740; MAM_2; 1.
DR PROSITE: PS00060; MAM; 1.
DR PROSITE: PS00024; SEA; 1.
DR PROSITE: PS00024; SEA; 1.
DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
DR PROSITE: PS00420; SRCR_2; 1.
DR PROSITE: PS00287; SRCR_2; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor;
KW Transmembrane; Zymogen.
FT CHAIN 1 829
FT CHAIN 830 1069
FT DOMAIN 1 18
FT DOMAIN 19 47
FT TRANSMEM 19 47
FT DOMAIN 48 1069
FT DOMAIN 52 169
FT DOMAIN 227 268
FT DOMAIN 270 379
FT DOMAIN 387 549
FT DOMAIN 569 679
FT DOMAIN 686 724
FT DOMAIN 723 816
FT DOMAIN 830 1069
FT ACT_SITE 874 874
FT ACT_SITE 925 925
FT ACT_SITE 1021 1021
FT DISULFID 229 242
FT DISULFID 236 255
FT DISULFID 249 266
FT DISULFID 688 700
FT DISULFID 695 713
FT DISULFID 707 722
FT DISULFID 817 945
FT DISULFID 859 875
FT DISULFID 959 1027

```

FT DISULFID 991 1006 By similarity.  
 FT DISULFID 1017 1045 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 147 147 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 197 197 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 212 212 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 373 373 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 380 380 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 433 433 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 515 515 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 579 579 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 675 675 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 727 727 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 751 751 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 770 770 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 791 791 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 897 897 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 936 936 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 999 999 N-linked (GlcNAc. . .) (Potential).  
 SQ SEQUENCE 1069 AA, 118735 MW, E62549B4637453CD CRC64;

Query Match 17.9%; Score 677.5; DB 1; Length 1069;  
 Best Local Similarity 29.2%; Pred. No. 4,9e-39;  
 Matches 186; Conservative 93; Mismatches 220; Indels 137; Gaps 23;

QY 86 LTRSPDLASCDERGSGLVTVYNTLSPEPHALVOLCGYPPSYNLTFFHSSQNVLLITLI 145  
 DB 522 VVFVAFNRRCSTALDIDISTNGICSGSPPEPTLVTPPP----- 563  
 QY 146 TINTERRRHGFATFPQLPRMSSCGG-RLRKAQSTFSPYRGHYPPNIDTWNIEVPPN 203  
 DB 564 -----ELPP--TDCGPELWEPNSTFSPNPDRTYPAQASCIMWLNARQ 606  
 QY 204 QHVAVRKFFYLLERPRACGTCFADYVEI--NGE-----KYCGESQFVTSNSKIT 254  
 DB 607 KNQGLHGEEDLEN-----INDVVEVADGGEFDSLALAVYGPPEVADLFTTNRMT 658  
 QY 255 VRHSDQSYTDGTGLAAYLS--YDSSDPC-PQGFCTRTGRCIKELRCGMADCTDHS 310  
 DB 659 VIETNNMTRRRKGFKNFTSGYVLGIPEPCODDEFQCKDNCLPLGNLCTSYPRCRDSD 718  
 QY 311 ELNCG-----SCDAHQFTCKNKKFKPLFWVDSVNDGSDNDEGCGCPAOTPRCSN 362  
 DB 719 EASGVRFNGTRSNNGLVQFNIHS-----IWHI--ACAEEN-----WTTQISN 758  
 QY 363 GKCLSKSQCGKXKDCGSDGDEASCP-----KVNVTCTKHYRCINGLCLSKGN 412  
 DB 759 EVC-----HLLGLGSANSMPLSSVCGGFVAVN-----QAPNGSLILTPS 799  
 QY 413 PECDGKE---DCSDGSDKDCGGLRFRQA--RVVGTDADGEMPMQVSLHALGQG 466  
 DB 800 LQCSODSLILLOCHNKS-----CGEKKVYQKVPKIVGSDAQAGAMPVVALYHRDS 853  
 QY 467 H---IGASLISPMWLSAAHCYIDDRGFYSDDTOMTAFGLHDSQORASAPQGERRLK 523  
 DB 854 TDRLLCASLIVSWLVAHSAHCYR---RLUDTTRTAVALGLHMGNLTPQVVRVVD 909  
 QY 524 RIISHPEFNDFTPDYALALELEKPAEYSSWVRPICPDASHVPAKKAIVWTGHTQY 583  
 DB 910 QIVINPIHYDRRKNDIAMMHELEKVAITYIQCIPBENQIFIPERTCSIAWGWDKI 969  
 QY 584 -GGTALLLOKEIRIVNOTTCENLPO-QITPRMCMVGLSGVDSQCGDSGGPLSSVE 641  
 DB 970 NAGSTVAVLKEADVPLISNEKCOQOLPEVYNTSEMICAGYEGGIDSCQDSGGPLMCOE 1029  
 QY 642 ADGRIPQAGVSWGDGCAQRNKPGVYTRLLPLFRMI 677  
 DB 1030 -NNRWFIVGVTSFVQCALPNHFGVYVRVSQFIEMI 1064

RESULT 29  
 CORI\_MOUSE STANDARD; PRT; 1113 AA.  
 AC Q92319;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-  
 DE converting enzyme) (Corin) (low density lipoprotein receptor related  
 DE protein 4).  
 GN Name=Corin; Synonyms=Crm, Lrp4;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98429596; PubMed=9756624;  
 RA Tomita Y., Kim D.-H., Magoori K., Fujino T., Yamamoto T.T.;  
 RT "A novel low-density lipoprotein receptor-related protein with type II  
 RT membrane protein-like structure is abundant in heart."  
 RL J. Biochem. 124:784-789 (1998).  
 CC -I- FUNCTION: Converts pro-ANP to ANP. Cleaves pro-ANP specifically  
 CC between Arg-122 and Ser-123 (By similarity).  
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein.  
 CC -I- TISSUE SPECIFICITY: Highly expressed in heart.  
 CC -I- SIMILARITY: Belongs to peptidase family S1.  
 CC -I- SIMILARITY: Contains 2 Fz-like (Fz) domains.  
 CC -I- SIMILARITY: Contains 7 LDL-receptor class A domains.  
 CC -I- SIMILARITY: Contains 1 SRCR domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC -----  
 CC EMBL; AB013874; BAA34371.1; -.  
 CC FTR; J03015; J03015.  
 CC HSRP; P01130; 1A0J.  
 CC MEROPS; S01.019; -.  
 CC MGD; MGI:1349451; Corin.  
 CC InterPro: IPR000024; Fz domain.  
 CC InterPro: IPR002172; LDL\_receptor\_A.  
 CC InterPro: IPR001254; Peptidase\_S1.  
 CC InterPro: IPR001314; Peptidase\_S1A.  
 CC InterPro: IPR009003; Pept\_Ser\_Cys.  
 CC InterPro: IPR001190; Srcr\_receptor.  
 CC Pfam; PF01392; Fz\_2.  
 CC Pfam; PF00057; Ldl\_recept\_a; 6.  
 CC Pfam; PF00069; Trypsin\_1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC PRINTS; PR00261; LDLRECEPTOR.  
 CC SMART; SM00063; FRI; 2.  
 CC SMART; SM00192; LDLa; 7.  
 CC SMART; SM00020; SR; 1.  
 CC SMART; SM00020; TRYP\_SPC; 1.  
 CC PROSITE; PS50038; Fz; 2.  
 CC PROSITE; PS01209; LDLRA\_1; 6.  
 CC PROSITE; PS50068; LDLRA\_2; 7.  
 CC PROSITE; PS00420; SRCR\_1; FALSE\_NEG.  
 CC PROSITE; PS50287; SRCR\_2; 1.  
 CC PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; FALSE\_NEG.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 CC Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor;  
 KW Transmembrane.  
 FT DOMAIN 1 112 Cytoplasmic (Potential).  
 FT TRANSMEM 113 133 Signal-anchor for type II membrane  
 FT protein (Potential).  
 FT DOMAIN 134 1113 Extracellular (Potential).  
 FT DOMAIN 134 327 Fz 1.  
 FT DOMAIN 336 372 LDL-receptor class A 1.  
 FT DOMAIN 373 408 LDL-receptor class A 2.  
 FT DOMAIN 409 445 LDL-receptor class A 3.

```

FT DOMAIN 446 483 LDL-receptor class A 4.
FT DOMAIN 518 548 Fz 2.
FT DOMAIN 647 682 LDL-receptor class A 5.
FT DOMAIN 683 721 LDL-receptor class A 6.
FT DOMAIN 722 757 LDL-receptor class A 7.
FT DOMAIN 758 853 SRCR.
FT ACT_SITE 869 1113 serine protease.
FT ACT_SITE 910 910 Charge relay system (By similarity).
FT ACT_SITE 959 959 Charge relay system (By similarity).
FT ACT_SITE 1052 1052 Charge relay system (By similarity).
FT DISULFID 857 979 By similarity.
FT DISULFID 895 911 By similarity.
FT DISULFID 1022 1037 By similarity.
FT DISULFID 1048 1077 By similarity.
FT CARBOHYD 147 147 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 202 202 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 208 208 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 298 298 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 317 317 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 373 373 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 411 411 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 444 444 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 481 481 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 519 519 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 537 537 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 635 635 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 719 719 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 765 765 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 828 828 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 970 970 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1089 1089 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 1113 AA; 122964 MW; B845B2C5F20DDEBC CRC64;

Query Match 17.5%; Score 663.5; DB 1; Length 1113;
Best Local Similarity 32.7%; Pred. No. 4,9e-38;
Matches 158; Conservative 67; Mismatches 187; Indels 71; Gaps 16;

QY 256 RFHSDGYTDFGLAEVLYSDSDPCPGFCRTGRCRKELRDGDMACTDHSDELNCS 315
DB 627 QPESSDNQCTLLP---NEVEECSPEHFRCRSGRCVLSGRCDGDQDDDEEBCG 683
QY 316 CDAGHQFTCK-NKFCCKPLFWVCDSDVNDGSDSDGSCS-CPAQTFRCSNGKLSKQOCN 373
DB 684 CKERALMECPNKKQCKHTLTCDFPRCDSDMERKNSFCQDNLEECANHECVRRDLCD 743
QY 374 GKDDCGSDGSDASCPRKN-----VTCYKTYRCLNG-----LC----- 407
DB 744 GWVDCSDSDDEMGCVTLSSKNGNSSLLTVHKSAXEHVCAQWRETLSQLACKQMGLEGP 803
QY 408 -LSKGNBECQKE-----DCSDGSDKDP-----C---DCGLASFT 438
DB 804 SVTLILFQEQEQWLRLYPNMENLNGSTLQELVYRHSCEPSRSEISLCSKQDGRPRPA 863
QY 439 R-QARVVGSTADAGCEWPMQVSLHALGQHICGASLISPNMLVSAHCYIDRGRFVSDP 497
DB 864 RMNRRIIGGRSRGRPRMPCSLQSBEGHICGCVLAKKVVLYAHCF-BGR-----EDA 918
QY 498 TQMTAFGLHDQSGRSPAGVGERLKRILISHPFNDFPDYDIALLELEAEVSSMVR 557
DB 919 DVWKVVFGINLIDHPSG-FMQTRFVKTLILHPRYSRAVVDYDISVELSDINETSYP 977
QY 558 ICLPDASHVFPAGKAIWVGHWGHTOYGTALIIQKGIKIRINOTGENTLLPQ-ITPRM 616
DB 978 VCLPSPEEYLEPDYCYITGMH--MGNKMPFKLQEGSVRLIIPLECCSYDMKTITRM 1035
QY 617 MCVFPLSGVSDSCGSDGSPLSVEADGRIFQAGVSVWGDCGAOR-NKPGVYTRLLPFRD 675
DB 1036 ICAIGESCTVSDSCMGSDGSPLVCEPRGQWTLFGLTMSGVSFSAKVLGPGVSVSYTVG 1095
QY 676 WIK 678
DB 1096 WIE 1098

```

```

RESULT 30
Q9DGR2 PRELIMINARY; PRT; 767 AA.
AC Q9DGR2;
DT 01-MAR-2001 (Tremblrel.16, Created)
DT 01-MAR-2001 (Tremblrel.16, Last sequence update)
DT 01-MAR-2004 (Tremblrel.26, Last annotation update)
DE Embryonic serine protease-2.
GN Name=xesp-2;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodidae; Xenopus.
NCBI_TaxID=8355;
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
RT from Xenopus laevis";
RL Gene 252:209-216 (2000).
CC -1-SIMILARITY: Belongs to peptidase family S1.
DR EMBL; AB038497; BAB08217.1; -.
DR HSSP; P00760; IEZX.
DR MEROPS; S01.049; -.
DR GO; GO:0016020; C:membrane, IEA.
DR GO; GO:0004263; F:chymotrypsin activity, IEA.
DR GO; GO:0008233; F:peptidase activity, IEA.
DR GO; GO:0005044; F:scavenger receptor activity, IEA.
DR GO; GO:0004295; F:trypsin activity, IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis, IEA.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF000059; Ldl_recept_a; 4.
DR Pfam; PF000089; Trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 8.
DR SMART; SM00020; TRYPSIN_1.
DR PROSITE; PS001209; LDLA_2; 8.
DR PROSITE; PS00068; LDLA_2; 2.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 767 AA; 86001 MW; E056A38796DE96E CRC64;

Query Match 17.0%; Score 644.5; DB 2; Length 767;
Best Local Similarity 26.6%; Pred. No. 7e-37;
Matches 170; Conservative 72; Mismatches 214; Indels 183; Gaps 17;

QY 193 DCTN-----LEVPNNQVKKRKFYLLERRACGTPKQVYENSKYC--GERSQ 243
DB 150 DCPYGDERNCAKTPPTPTQWCSYY-----TCYVQICNGVDCYGDDE 200
QY 244 FVVTNSNKTIVFPHSQSYTDGFLAEVLYSDSDPCPGFCRTG----- 290
DB 201 NCATKTSITPCQWYC--STSTCIYGYQ-----CNGVQCAVDDERNCAKTPSI 251
QY 291 -----RCIRKELRCDGMACTDHSDELNCSGDAGHQFTCK-----NKFCKP 331
DB 252 PTCQLYSSYYTCTIAYQICNGVLDCEPVDDERNCAVATTSPTCQICWDPMEDYTCIY 311
QY 332 LFWVCDSDVNDGSDSDGSC-----SCPAQTFRNS-----CKLSKSGQCNCKD 376
DB 312 AYQWCDGVRQCYGDDEKNCVATTTATTSPPTCQIYCMNPMYYTCTIAYQWCDGVR 371
QY 377 DCGDSDGASC-PRVNVVVTCTKATYRCLNGCLSKGNPECDGKEDCSGSDKDC----- 430

```

```

Db      372 QCYGDBELNCDTRTTAYCEK---RCSSSVSCVLSQWCDGVSDCPYGEDMSCVSLYP 428
QY      431 -----430
Db      429 ADFOLOVYSTSVSAMLFCSDYMNDDGFRACODFGYNGSSYNRYDTLMSFYAPNGYFKL 488
QY      431 -----DCGLRSFTRQARVVGSTDADEGEWPMQVSL 460
Db      489 YSGYWRSEKFTYSVQYSSYCYSGNVVSLHCTSCGVSNNSLVSRIVCGTFANLGNWPMQVNL 548
QY      461 HALGQGHICGASLISPMWLVSAHCTIDDKGFRYSDEPTQWTAFLGLHDQSORSAFVQER 520
Db      549 QYI-TGVLGSGSIISPMWIVTAHCVYGS---YSSASGWRVFAG--TLTKPSYYNASAY 601
QY      521 RLKRIISHPPFNDFPYDIALALELEKPAEYSSMWRPICLPDASHVPFAGKAIWVTGMH 580
Db      602 FVERIIVHPGKSYTYNDIALMKLRDEITFGYTTQVPCLPNSGMFWBAGTTTWISGMS 661
QY      581 TOYGTGALILQKEIRVINQTCEN--LLPOQITPRMFCVFLSGGVDSQGDSDGGPLS 638
Db      662 TYEGGSYSTYLQYVAIPLIDSNVCNQSYYVNGQITSSMICAGYLSGGVDTQGDSDGGPLV 721
QY      639 SVEADGRIFOAGVVSQDGCQQRNKPQVYTRLPFLFRDVI 677
Db      722 N-KRNGTWMVLVGDTSWGDGCARANKPGVYGNVTFLEWI 759

```

Search completed: November 29, 2004, 08:32:23  
 Job time : 123.685 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: November 29, 2004, 08:19:58 ; Search time 25.3127 Seconds

(without alignments)  
2596.165 Million cell updates/sec

Title: US-09-936-333-5  
Perfect score: 3789  
Sequence: 1 MAERVYVLPFRARSLKSFV.....PGVYTRLPFRMDIKENTGV 683

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : PIR-79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3197	84.4	855	2 JCT7731	membrane-bound arg
2	718.5	19.0	1034	1 A53663	enteropeptidase (E
3	703.5	18.6	1035	1 A43090	enteropeptidase (E
4	683.5	18.0	1019	1 A56318	enteropeptidase (E
5	663.5	17.5	1113	2 JE0315	low-density lipopr
6	576.5	15.2	1524	2 T30337	polyporein - Afri
7	565	14.9	699	1 I54763	Ra-reactive factor
8	545.5	14.4	638	1 KQMSPL	plasma kallikrein
9	526.5	13.9	638	1 KQRTPL	plasma kallikrein
10	506	13.4	613	2 S15468	complement C3b/C4b
11	504	13.3	790	1 B1545	plasmin (EC 3.4.21
12	503	13.3	460	2 B61545	plasmin (EC 3.4.21
13	501.5	13.2	786	1 A47547	serine proteinase
14	499	13.2	638	1 KQHUP	plasma kallikrein
15	492.5	13.0	583	1 A29154	complement factor
16	492.5	13.0	810	1 PLHU	plasmin (EC 3.4.21
17	491.5	13.0	812	1 PLMS	plasmin (EC 3.4.21
18	491	13.0	416	1 KFB0	coagulation factor
19	490	12.9	417	1 S00845	hepsin (EC 3.4.21.
20	484.5	12.8	343	1 A57014	protease (EC 3.4.
21	484.5	12.8	810	1 KFHU	coagulation factor
22	483	12.7	810	2 I46260	plasmin (EC 3.4.21
23	481.5	12.7	812	1 PLBO	plasmin (EC 3.4.21
24	481	12.7	810	2 B30848	plasmin (EC 3.4.21
25	480.5	12.7	416	1 S33777	hepsin (EC 3.4.21.
26	480	12.7	625	1 KFHU1	coagulation factor
27	474.5	12.5	452	1 A50351	coagulation factor
28	474	12.5	686	1 A59271	Ra-reactive factor
29	472	12.5	275	2 S40007	trypsin (EC 3.4.21

30	472	12.5	455	2 A61545	plasmin (EC 3.4.21
31	468.5	12.4	2616	2 A57096	nudel protein prec
32	467	12.3	275	2 A50005	trypsin (EC 3.4.21
33	467	12.3	562	1 UKHTT	t-plasminogen acti
34	462	12.2	274	2 S35339	trypsin (EC 3.4.21
35	461.5	12.2	267	2 S40006	trypsin (EC 3.4.21
36	460	12.1	1420	2 A32869	apolipoprotein(a)
37	452.5	11.9	560	1 JC4795	plasma hyaluronan
38	451	11.9	238	1 TRWVSJ	trypsin-like prote
39	449.5	11.9	4548	1 S00557	apoptotain(a) (EC
40	442.5	11.7	1004	2 T30338	oviductin (EC 3.4.
41	441	11.6	254	1 TRWV3Y	trypsin-like prote
42	439	11.6	270	2 S56160	mast cell trypase
43	438.5	11.6	459	2 JQ0419	coagulation factor
44	437.5	11.5	247	1 TRDC	trypsin (EC 3.4.21
45	436	11.5	242	2 S49489	trypsin (EC 3.4.21
46	436	11.5	263	2 A21195	chymotrypsin (EC 3
47	434	11.5	263	1 KYRTB	chymotrypsin (EC 3
48	434	11.5	265	2 T15451	hypothetical prote
49	433.5	11.4	273	2 A47246	trypase (EC 3.4.2
50	433	11.4	273	2 S35340	trypsin (EC 3.4.21
51	431.5	11.4	231	1 TRPGTR	trypsin (EC 3.4.21
52	431	11.4	276	2 A38654	mast cell proteina
53	428.5	11.3	247	1 B25852	trypsin (EC 3.4.21
54	427.5	11.3	247	2 A27547	trypsin (EC 3.4.21
55	427.5	11.3	622	1 TBHU	thrombin (EC 3.4.2
56	426.5	11.3	407	1 KFB07	coagulation factor
57	426	11.2	231	2 S31778	trypsin (EC 3.4.21
58	425.5	11.2	558	2 JC6878	plasma hyaluronan
59	425	11.2	275	2 A32410	trypase (EC 3.4.2
60	424	11.2	274	2 JC4171	trypase (EC 3.4.2
61	422	11.1	263	2 A31299	chymotrypsin (EC 3
62	422	11.1	275	2 B35863	trypase (EC 3.4.2
63	421	11.1	246	2 B25528	trypsin (EC 3.4.21
64	421	11.1	246	2 JQ1472	trypsin (EC 3.4.21
65	421	11.1	275	2 A35863	trypase (EC 3.4.2
66	420.5	11.1	266	2 S54166	trypsin (EC 3.4.21
67	420.5	11.1	304	2 S33496	trypsin (EC 3.4.21
68	420.5	11.1	559	1 A35029	t-plasminogen acti
69	420.5	11.1	705	1 C1HURB	complement subcomp
70	420	11.1	245	1 KYBOB	chymotrypsin (EC 3
71	419.5	11.1	559	1 A29941	t-plasminogen acti
72	419.5	11.1	761	2 JC5759	brain-specific ser
73	419	11.1	246	2 JQ1471	trypsin (EC 3.4.21
74	418.5	11.0	456	1 KXBO	protein C (activat
75	418	11.0	229	1 TRDPS	trypsin (EC 3.4.21
76	418	11.0	618	2 A35827	thrombin (EC 3.4.2
77	417.5	11.0	246	1 TRRT1	trypsin (EC 3.4.21
78	417	11.0	242	2 S31776	trypsin (EC 3.4.21
79	417	11.0	242	2 S31775	trypsin (EC 3.4.21
80	417	11.0	275	2 C35863	trypase (EC 3.4.2
81	417	11.0	462	1 EXRT	coagulation factor
82	415.5	11.0	247	2 S13813	trypsin (EC 3.4.21
83	415	11.0	247	2 S05494	trypsin (EC 3.4.21
84	415	11.0	432	1 S18932	u-plasminogen acti
85	413.5	10.9	274	2 A45754	trypase (EC 3.4.2
86	412.5	10.9	238	2 S31779	trypsin (EC 3.4.21
87	412.5	10.9	431	2 S47538	acrosin (EC 3.4.21
88	412	10.9	263	2 S47537	chymotrypsin (EC 3
89	411.5	10.9	250	2 S55493	serine proteinase
90	411	10.8	247	1 A25852	trypsin (EC 3.4.21
91	410	10.8	266	2 JC4850	trypsin-like prote
92	409.5	10.8	237	1 TRCV1	trypsin (EC 3.4.21
93	409.5	10.8	274	2 I47078	coagulation factor
94	408	10.8	245	1 KYBOA	chymotrypsin (EC 3
95	408	10.8	259	2 I38363	trypsin (EC 3.4.21
96	407.5	10.8	264	2 I38136	chymotrypsin-like
97	407	10.7	492	1 EXBO	coagulation factor
98	406.5	10.7	246	1 TRDGC	trypsin (EC 3.4.21
99	406.5	10.7	466	1 KFHU7	coagulation factor
100	406.5	10.7	617	2 S10511	thrombin (EC 3.4.2

## ALIGNMENTS

## RESULT 1

JC7731

membrane-bound arginine-specific serine proteinase precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C:Accession: JC7731; JC7775

R:Kish, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Achauda, J. Biochem. 130, 425-430, 2001

A:Title: Characterization of a membrane-bound arginine-specific serine protease from rat

A:Reference number: JC7731; MUID:21421307; PMID:11530019

A:Accession: JC7731

A:Molecule type: mRNA

A:Residues: 1-855 &lt;RTS&gt;

A:Cross-references: UNIPROT:Q9JUI7; DDBJ:AB049189

A:Experimental source: strain Male, 7-week-old

R:Satom, S.; Yamazaki, Y.; Tsuzuki, S.; Hicomi, Y.; Iwanaga, T.; Fushiki, T. Biochem. Biophys. Res. Commun. 287, 995-1002, 2001

A:Title: A role for membrane-type serine protease (MT-SPL) in intestinal epithelial turn

A:Reference number: JC7775; PMID:11573963

A:Contents: Small intestine

A:Accession: JC7775

A:Molecule type: mRNA

A:Residues: 1-855 &lt;SAT&gt;

A:Cross-references: DDBJ:AB037898

C:Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease of specific proteins or peptides on the brushborder membranes. It also participates in lial migration and/or cell loss.

C:Genetics:

A:Gene: mt-spl

A:Map position: basolateral cell surface

C:Superfamily: membrane-bound arginine-specific serine proteinase

C:Keywords: protein digestion

Query Match 84.4%; Score 3197; DB 2; Length 855;

Best Local Similarity 83.0%; Pred. No. 3.7e-203;

Matches 567; Conservative 59; Mismatches 57; Indels 0; Gaps 0;

```
QY 1 MAERVVMLPPRAASLSFVVTSVVAPPTDSKYQRTQDNSSCSGLHARGVELMRFTTPG 60
DB 173 MAVRRVVTLPRAALSKFVLTVAAPDIPRMLOKRTQDNSSCSGLHARGVTRFTTPG 232
QY 61 PPDSPPVAHARCOMALRGDADSVLTFPSFDLASCSERGSDDLVTYVNTLSMPMFALVQ 120
DB 233 FPNPSYPAHARCOMALRGDADSVLTFPSFDVA PCGHSDLVTVDSLSMPMFALVVR 292
QY 121 LCGTYPSPYMLTFHSGQNVLLITLITERRHHPGFATFFQLPRMSSCGRLRKAQGTEN 180
DB 293 LCGTSPSPYMLTFHSGQNVLLITLITERRHHPGFATFFQLPRMSSCGRLRKAQGTFS 352
QY 181 SPYRGGHYPRNIDCTNMIIEVPPNNQVVRKRFLLPRACGCPDYVEINERKCGE 240
DB 353 SPYRGGHYPRNIDCTNMIIEVPPNNQVVRKRFLLPRACGCPDYVEINERKCGE 412
QY 241 RSQFVTVSNKTIIVRHSDOSYTDGTFLAEYLSDSDSCPOGFTCRGCIKELKRC 300
DB 413 RSQFVTVSNKTIIVRHSDOSYTDGTFLAEYLSDSDSCPOGFTCRGCIKELKRC 472
QY 301 GMADCTHSDDELNCSCDAHQFTCKNFKCPLFMVCDSDVDCGSDNDEGSCCPAQTRC 360
DB 473 GMADCPYSDERHRCNATHQFMCKNFKCPLFMVCDSDVDCGSDNDEGSCCPAQTRC 532
QY 361 SNGKCLSKSQOCCNGKDCGSDGSDASCPKVVNTCTGTRCLNGLCLSKNPFCDKED 420
DB 533 SNGKCLSKSQOCCNGKDCGSDGSDASCPKVVNTCTGTRCLNGLCLSKNPFCDKED 592
QY 421 CSDDSDKDCDCCGLRFTROARVVGGTDADGEGWPMOVSLHALGQGHICGASLSPMLV 480
DB 593 CSDDSDKDCDCCGLRFTROARVVGGTDADGEGWPMOVSLHALGQGHICGASLSPMLV 652
QY 481 SAACVYIDRGFRYSDEPTQWTAFLGLHDQGRSAPGVQERRLKRIISHPFENDFTPDYI 540
```

```
DB 653 SAACHFODETIFKXSDHTWMTAFGLHDQSKRSASGVQEHKLRITTHSFNFTPDYDI 712
QY 541 ALLEERKPAEYSSMNPICLPDASHVFPAGKATWGMQHTGVGGALLILOGEIRVYN 600
DB 713 ALLEERKPAEYSSMNPICLPDNTWHPAGKATWGMQHTGVGGALLILOGEIRVYN 772
QY 601 QTTCEMLLPQILPRMNCVGLSGVDSCQSDGSGPLSVYADGRIFQAGVWSMGSCAQ 660
DB 773 QTTCEMLLPQILPRMNCVGLSGVDSCQSDGSGPLSVYADGRIFQAGVWSMGSCAQ 832
QY 661 RNKPGVYTRLPFRDWIKENTGV 683
DB 833 RNKPGVYTRLPFRDWIKENTGV 855
```

## RESULT 2

A53663

enteropeptidase (EC 3.4.21.9) precursor [validated] - pig

N:Alternate names: enterokinase

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 07-Oct-1994 #sequence\_revision 09-Aug-1996 #text\_change 28-Apr-2003

C:Accession: A53663

R:Matsumura, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa, J. Biol. Chem. 269, 19976-19982, 1994

A:Title: Structural characterization of porcine enteropeptidase.

A:Reference number: A53663; MUID:94327548; PMID:8051081

A:Accession: A53663

A:Molecule type: mRNA

A:Residues: 1-1034 &lt;MAT&gt;

A:Cross-references: GB:D30799; NID:G505122; PID:BA06459.1; PID:G505123

A:Note: parts of this sequence, including the amino ends of three chains isolated from ct

C:Comment: The mechanism of association with the membrane of the intestinal brush border

C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)

ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involve

C:Function:

A:Description: cleaves activation peptide from trypsinogen to produce active trypsin

A:Pathway: intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding repe

C:Keywords: glycoprotein; hydrolase; serine proteinase; LDL receptor ligand-binding repe

F:22-33/Domain: transmembrane #status predicted &lt;TM&gt;

F:52-117/Product: enteropeptidase mini chain #status predicted &lt;MCH&gt;

F:118-799/Product: enteropeptidase heavy chain #status predicted &lt;HCH&gt;

F:199-236/Domain: LDL receptor ligand-binding repeat homology &lt;LDL1&gt;

F:357-519/Domain: MAM homology &lt;MAM&gt;

F:511-646/Domain: Clr/Cls repeat homology &lt;Clr&gt;

F:658-692/Domain: LDL receptor ligand-binding repeat homology &lt;LDL2&gt;

F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical &lt;SRCr

F:800-1034/Product: enteropeptidase light chain #status predicted &lt;LCH&gt;

F:800-1034/Domain: trypsin homology &lt;TR&gt;

F:116-147, 170, 194, 283, 343, 350, 403, 455, 485, 518, 549, 645, 697, 701, 721, 740, 761, 804, 863, 902, 964

F:787-911, 825-841, 925-992, 956-971, 982-1010/Disulfide bonds: #status predicted

F:840-891, 986/Active site: His, Asp, Ser #status predicted

Query Match 19.0%; Score 718.5; DB 1; Length 1034;

Best Local Similarity 32.4%; Pred. No. 1e-39;

Matches 181; Conservative 87; Mismatches 189; Indels 101; Gaps 22;

```
QY 161 QLPKSSCGG--RLRAQGTNSPYRPPNIDCTNMIIEVPPNNQVVRKRFLLPRACGCPDYVEINERKCGE 218
DB 534 ELP--TDCCGFPELMEPNTFTSMNPNNYPNDAFCVMNINQKGNIDLHPEEFDLEN- 590
QY 219 RRAQGTCPDYVEI--NGEK-----YQERSQFVTVSNKTIIVRHSDOSYTDGTFL 269
DB 591 -----LADVVEIRGEEEDSLLAAYVTPGPEVEYFTTNMTVLFTINDLITGSGFR 643
QY 270 AEYLS---YDSDPDC--PGQFTCRGRCIRKELKCDGMADCTHSDDELNCSCDAHQFTCK 325
DB 644 ANFTTGYHIGIRPCEDNPFQGENGECVLLVNCDSFHSXKQSDAHNCV----- 693
QY 326 NFKCPLFMVCDSDVDCGSDNDEGSCCPAQTRFRC--NGKCLSKSQOCCNGKDC---G 379
```



Db 694 -RF-----LNGTANSG-----LVQFRIQSIWHTACAEWTTQTSDDVCQLLG 735  
 Qy 380 DGSBASC-----KYNVVTCTKHTKRLNGCLSLSGNPECDKE-----CSD 423  
 Db 736 LGTGNSSHPFSSGGGPPVKLNTAP-----NGSLILPASECC--FESLLILQCNH 784  
 Qy 424 GSDKDCDCGRLSFTQA--RVVGSTDADEGEMPVQVSLHALGQHICGASLISPMVLVS 481  
 Db 785 KS-----CGKQVAYQEVSPKIVGDNDRBEGAMPVVALYVNGQ--LLCGASLVSMDWLS 837  
 Qy 482 AARCYIDRRGRYSDDPTQWTAFLGLHDQSQRSAPEVGRRLKRIISHPFNDFTFDVIA 541  
 Db 838 AARCYVYG---RNLPESSKMAKILGLHMTSNLTSPOIVTRLIDELIVNPHYRRRDXSDIA 893  
 Qy 542 LLELEKPAEYSSMRPCLPDASHVFPAGKAIWVGKHTQYGTGALLQGEIRVING 601  
 Db 894 MHLPEKVNVTYDIQICLPEBNQVFPGRICSLAGMKVLYQSPADIIQEDADPPLLSN 953  
 Qy 602 TTCENLTPQ--QITPRMVCVFLSGVDSGQDSGGLSSVEADGRIFQAGVSMGCGAQ 660  
 Db 954 EKCQGMPEYNTENNMKAGYBEGIDBSCQDSGGLMCLF--NNRWLAGVTSFGYQCAL 1012  
 Qy 661 RNKPGVYTRLPFLFDMIK 678  
 Db 1013 PNRGVYARVPRFTTEWIG 1030

RESULT 3  
 A43090  
 enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine  
 N/Alternate names: enterokinase  
 C/Species: Bos primigenius taurus (cattle)  
 C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004  
 C/Accession: A43090; A48874; A61436  
 R/Klamato, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994  
 A/Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo  
 A/Reference number: A43090; MUID:94329561; PMID:8052624  
 A/Accession: A43090  
 A/Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1035 <KIT>  
 A/Cross-references: UNIPROT:P98072; GB:U09859; NID:9746410; PID:AA840026.1; PID:9746411  
 A/Experimental source: small intestine  
 R/Lavallie, E.R.; Rehmetulla, A.; Racle, L.A.; DiBlasio, E.A.; Perez, C.; Grant, K.L.;  
 J. Biol. Chem. 269, 23311-23317, 1993  
 A/Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of h  
 A/Reference number: A48874; MUID:94043122; PMID:8226855  
 A/Accession: A48874  
 A/Molecule type: mRNA  
 A/Residues: 801-1035 <LAV>  
 A/Cross-references: GB:L19663; NID:9416131; PID:AAA16035.1; PID:9416132  
 A/Note: parts of this sequence, including the amino end of the mature protein, were conf  
 R/Light, A.; Janska, H.  
 J. Protein Chem. 10, 475-480, 1991  
 A/Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.  
 A/Reference number: A61436; MUID:92189715; PMID:1799406  
 A/Accession: A61436  
 A/Molecule type: protein  
 A/Residues: 801-807, 'Y', 809-827 <LIG>  
 C/Comment: The mechanism of association with the membrane of the intestinal brush border  
 C/Comment: Conversion from membrane-bound to soluble forms may involve further processin  
 C/Comment: mature enteropeptidase is variously reported to contain two (heavy and light)  
 l fide linked  
 C/Function:  
 A/Description: cleaves propeptide from trypsinogen to produce active trypsin  
 A/Pathway: intestinal digestive hydrolase cascade  
 C/Superfamily: enteropeptidase, C1r/C1s repeat homology; LDL receptor ligand-binding reg  
 C/Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein  
 F/22-38/Domain: transmembrane #status predicted <TM>  
 F/52-117/Product: enteropeptidase mini chain #status predicted <MCH>  
 F/118-800/Product: enteropeptidase heavy chain #status predicted <HCH>

F/199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
 F/358-520/Domain: MAM homology <MAM>  
 F/542-647/Domain: C1r/C1s repeat homology <C1R>  
 F/653-653/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
 F/694-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRCI  
 F/801-1035/Product: enteropeptidase light chain #status predicted <LCH>  
 F/801-1030/Domain: trypsin homology <TRY>  
 F/116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Binding  
 F/788-912,826-842,926-993,957-972,983-1011/Disulfide bonds: #status predicted  
 F/841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 18.6%; Score 703.5; DB 1; Length 1035;  
 Best Local Similarity 32.6%; Pred. No. 1e-38;  
 Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;

Qy 161 QLPMSCCG--RRLKAGFPNSPYRGHYPNIDCTWNIENVNQNVRKRFYLLP 218  
 Db 535 ELP--TDCGPHDIMEENTFTSINFENSPNOAFCITWNLNAQKXNIOHFOEDLEN- 591  
 Qy 219 RRAGCTPKDYVEI--NGEKYCGERSQFV-----VTSNKKITVRPHSDSYTDT 266  
 Db 592 -----IADVETRDEE--GDSLFLLAYTGGPVNDVFSITNRKTVLFTDNMLAKQ 641  
 Qy 267 GLFLAEYLS---YDSSDPC-PQOPTCRTRGCRKELKCDGADCTDSDELNC-----SC 316  
 Db 642 GFKANFTTGYGLGIPEDCKEDNFCQKDECIPLVNLCDGPFHCKDGSDEAHCVRLFNGTT 701  
 Qy 317 DAGH--QFTCKNKKCKLFWVCDSDVNDGSDNSDEQSCSPAQTRFGSNGKLSQCCNG 374  
 Db 702 DSGELVQFRIOSS-----IMHV-----ACAEN-----MTQISDDVC-----QLLG 736  
 Qy 375 KDDGSDSDSEASC-----PRVNVVCTKHTKRLNGCLSLSGNPECDKEGSDGDEK 428  
 Db 737 ---LGTNNSVPTRTSGGPPVNLNTAP-----NSLILTPGQC-----LEDLIL 781  
 Qy 429 DCD---CGLSFTQA--RVVGSTDADEGEMPVQVSLHALGQHICGASLISPMVLVSA 483  
 Db 782 QCNVKSQGLKVLVQEVSPKIVGSDSRGAMPVVALYPDQ--QVCASLVSMDWLSAA 840  
 Qy 484 HCYIDDRGRYSDDPTQWTAFLGLHDQSQRSAPEVGRRLKRIISHPFNDFTFDIAL 543  
 Db 841 HCVVG---RNLPESSKMAKILGLHMTSNLTSPOIETRLIDQIVNPHYRRRKNNDIAM 896  
 Qy 544 ELEKPAEYSSMRPCLPDASHVFPAGKAIWVGKHTQYGTGALLQGEIRVING 603  
 Db 897 HLEKVNVTYDIQICLPEBNQVFPGRICSLAGMKVLYQSPADVLEADVPPLLSN 956  
 Qy 604 CENLTPQ--QITPRMVCVFLSGVDSGQDSGGLSSVEADGRIFQAGVSMGCGAQ 662  
 Db 957 CQQGMPEYNTENNMKAGYBEGIDBSCQDSGGLMCLF--NNRWLAGVTSFGYQCALPN 1015  
 Qy 663 KPGVYTRLPFLFDMIK 678  
 Db 1016 PNRGVYARVPRFTTEWIG 1031

RESULT 4  
 A56318  
 enteropeptidase (EC 3.4.21.9) precursor [validated] - human  
 N/Alternate names: enterokinase  
 C/Species: Homo sapiens (man)  
 C/Date: 19-May-1995 #sequence revision 09-Aug-1996 #text\_change 09-Jul-2004  
 C/Accession: A56318; B43090  
 R/Klamato, Y.; Veille, R.A.; Donie-Keller, H.; Sadler, J.E.  
 Biochemistry 34, 4562-4568, 1995  
 A/Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolyt  
 A/Reference number: A56318; MUID:95234679; PMID:7718557  
 A/Accession: A56318  
 A/Molecule type: mRNA  
 A/Residues: 1-1019 <KIT>  
 A/Cross-references: UNIPROT:P98073; GB:U09860; NID:9746412; PID:AAAC50138.1; PID:9746413  
 R/Klamato, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

**A>Title:** Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of  
**A:Reference number:** A43090; MUID:94329561; PMID:8052624  
**A:Accession:** B43090  
**A:Status:** nucleic acid sequence not shown  
**A:Molecule type:** mRNA  
**A:Residues:** 749-1019 <K12>  
**A:Cross-references:** GB:U09860  
**C:Comment:** The mechanism of association with the membrane of the intestinal brush border (created below) or with amino-terminal myristoylation of the heavy chain.  
**C:Genetics:**  
**A:Gene:** GDB:PRSS7  
**A:Cross-references:** GDB:384083; OMIM:226200  
**A:Map position:** 21q21-21q21  
**C:Complex:** Mature enteropeptidase is variously reported to contain two (heavy and light) subunits by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involves cleavage.  
**C:Function:**  
**A:Description:** cleaves activation peptide from trypsinogen to produce active trypsin  
**A:Pathway:** Intestinal digestive hydrolyase cascade  
**C:Superfamily:** enteropeptidase; C1s/C1s repeat homology; LDL receptor ligand-binding repeat  
**C:Keywords:** glycoprotein; hydrolyase; serine proteinase; transmembrane protein; zymogen  
**F:1-784/Product:** enteropeptidase heavy chain #status predicted <HCH>  
**F:22-38/Domain:** transmembrane #status predicted <TM>  
**F:168-221/Domain:** LDL receptor ligand-binding repeat homology <LDL1>  
**F:342-504/Domain:** MAM homology <MAM>  
**F:526-631/Domain:** C1r/C1s repeat homology <C1r>  
**F:643-677/Domain:** LDL receptor ligand-binding repeat homology <LDL2>  
**F:678-783/Domain:** scavenger receptor cysteine-rich domain homology #status atypical <SRC>  
**F:785-1019/Product:** enteropeptidase light chain #status predicted <LCH>  
**F:785-1014/Domain:** trypsin homology <TRY>  
**F:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site:** F:178-886,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted  
**F:825,876,971/Active site:** His, Asp, Ser #status predicted

**Query Match** 18.0%; Score 683.5; DB 1; Length 1019;  
**Best Local Similarity** 32.5%; Pred. No.2,1e-37;  
**Matches** 180; Conservative 80; Mismatches 202; Indels 91; Gaps 20;

Oy	161	QLPMSSCGG--RLRKAOGTNSPYRGHPYRNIDCTWNIEVNNQHVKKFKFYLLP	218
Db	519	ELP--TDCGGPEELMEPEPTTFSSTNFPNSYNLAFCWILAAQKNIQLHFEDELEN	575
Oy	219	RRAGCTCKDVEI--NGEK-----YCGRSQFVWTNSNKTITVPHSQSTDTGFL	269
Db	576	-----INDVVELRDEADSLAVYTGCGPYKVDFTSTNRKTVLLITNDVLRAGCFK	628
Oy	270	AEYLS---YDSSDPQPG-QFTCTGRCLRKELRCGDWADCTDHSDELNCSGDAQHF	322
Db	629	ANFTTGYHLGIPRECKADHPQCKNGECVPLVNLCDGHLCEGSDGDEADCV-----RPFNG	683
Oy	323	TCKR----KFKCKPLFWVCDSDVNDGDNSEDEGSCSPAOTFPCSNKGKCLSKSGCNGDDC	378
Db	684	TTNNNGLVRFRIQSIW--HTCAEN-----WTTQISNDVC-----QLL	719
Oy	379	GDGSDDEASCPKVNVTCTKHTYRCINGLCSKNGPECDS-----KEDCSDGS-----D	426
Db	720	GLGSGNSKPIFST-----DGGPFKNTAPRPHGHLILPSCQCLDLSILRQCN	768
Oy	427	EKDCDCGLRSTRQARVVGTDADBGEMPMQVSLHALGQGHICGASLISPMWLVSAAHCY	486
Db	769	HKSCKKTLAADIPTKIVYGGNSNAKEGMPWVVGILY-YGGRLLCGASIVSSPMWLVSAAHCV	827
Oy	487	IDDGPRYSDDTQWAPFLGLHDOSORSAPOGEBRLKRIISHPRFNPTPTPYDILLLE	546
Db	828	YG---RNLEBSKTAIILGLMKSNLTSPOVPRPLIDETIVNPHYRRRRKNDIAPMMLLE	883
Oy	547	KPAEYSSWARDICLEDASHVEPAGKAIWVTWGMHTQYGGTGALLLQKGEIRVINQTCEN	606
Db	884	EKNVYTDVIGICLPEENQVFPFPGNCSIAQMGTVVVGTTANITLQADVPLLSNERCQ	943
Oy	607	LLPQ-QITPRMVCYGLSGSYDSCQSDSGGSLSSVEADGRIFQAGVYSSWEGDCGQARKKPG	665
Db	944	QMPENITENNIICAGYEEGGIDSCQSDSGGLMCE--NNRFFLAGVYSPFGKCALPFRPG	1002

[illegible]

T30337  
polyprotein - African clawed frog  
A:Title: Xenopus laevis (African clawed frog)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T30337  
R:Yang, J.C.; Lindsey, L.L.; Hedrick, J.L.  
submitted to the EMBL Data Library, March 1998  
A:Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from X.  
A:Reference number: Z20829  
A:Accession: T30337  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1524 <YAN>  
A:Cross-references: UNIPROT:Q91674; EMBL:U81290; NID:92981640; PID:92981641; PIDN:AA0247  
C:Superfamily: tryosin related polypeptide; trypsin homology

Query Match 15.2%; Score 576.5; DB 2; Length 1524;  
Best Local Similarity 25.9%; Pred. No. 3.5e-30;  
Matches 165; Conservative 91; Mismatches 206; Indels 175; Gaps 22;

QY 66 YPARARQWALRGADSVLSLFRFDL--ASCDEKSDLVVYNTLS-EMERHALVOL 121  
DB 333 YSINSVCWMLAVOKATIEIRFLQDIEDHATC-----FDYLSFTVNEKMKIRK 383  
QY 122 CGTVPSPYNTLTFHSSQVNLITLITNTERRRHGFEPATFPOLP--RMSGCG-GRLEKAGGT 178  
DB 384 CGSTIIPSLIY---RSNKVYTFPSDGTFTGREGFIQFLAIPTRKASACGSAKILIKKGM 440  
QY 179 FNSPYPHGYPNIDCTNIEVNNQHVAFKPFYLLPFRACGTCPKYVEI-NG--- 234  
DB 441 IYSNYPDPYRLTKCSWIIEAPENHIVKLFEDENV---EYGHGCIYDAVEYDGAEE 496  
QY 235 ----EKYGGESQPVVNSNKITVRHFSDDSYDTGGLAYLSVSDSPERGQFTCKTG 290  
DB 497 KQIARLCGYTLPLPISSPENTMILRFXTDENSYPGFKVAF----- 538  
QY 291 RCIRKELRCQGMADCTHSDDELNCSDAQHQTCKNK-FCKPLFWVCDVNDGDNDEQ 349  
DB 539 ----SFKREKQFSLPV----- 551  
QY 350 GCSGPAQTFRCNSGKSLKSOQCGKDDCGDGSDEASCPKXNVVYTCRHYRCLNGCLSL 409  
DB 552 ----DD-----TPTISML----- 560  
QY 410 KGNPECCKEDKSDSKDDCCGLRSTRTQ---ARVVGCTDADBGEMPMQVSLHALQCG 466  
DB 561 --HPRALAD-----VCGMAPMTPKMWLPRIYVGESEASPSNPWVOIIFPLRTF 607  
QY 467 HICGSLISPMVLVAHCYIDRGRFYSDPTOMTAFGLHDQSORSAPGVQERLKRIL 526  
DB 608 H-CGGAISIPWILTAHC-----IRAAESYMTVINGDNRMLNEST-EQIRNIKITR 659  
QY 527 SHPFNDFTFDYIALLELEKPAEYSSMWRPICLPDASHVPAGKAIWVTMGHTOYCGT 586  
DB 660 IHDVNSSETYNDIALALTYLEBPFLDLDNFRVCLPEPEPREVLTPASVCVVWTGNTAEDQG 719  
QY 587 GALLQGEIIVNQTGCE-NLLPQGITPRMCMGFLSG-GVDSQGGSGGFLSSVADG 644  
DB 720 PALGIQQLPLRLISICTNTSYSGELTDHMLCAGFFSSKSKDACCQDGSGLVCONEKE 779  
QY 645 RIFQAGVVMGDCGQNRKPGVYTRLPPLFRDMIKENT 681  
DB 780 QFSTYGLVSWGEGCGRVSKPGVYTKVRLFPFWI-ONT 815

RESULT 7  
154763  
Ra-reactive factor (BC 3.4.21.-) 1 precursor - human  
N:Alternate names: mannose binding protein-associated serine proteinase 1 (MASP-1)  
C:Species: Homo sapiens (man)  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C:Accession: I54763; JN0883  
R:Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.

Int. Immunol. 6, 665-669, 1994  
A:Title: Molecular characterization of a novel serine protease involved in activation of  
A:Reference number: I54763; MUID:94289349; PMID:8018603  
A:Accession: I54763  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-699 <SAT>  
A:Cross-references: UNIPROT:P48740; GB:D28593; NID:9790963; PIDN:BA05928.1; PID:9471128  
R:Takada, P.; Takayama, Y.; Hatanoue, H.; Kanakami, M.  
Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993  
A:Title: A new member of the C1s family of complement proteins found in a bactericidal f  
A:Reference number: JN0883; MUID:94059062; PMID:8240317  
A:Accession: JN0883  
A:Molecule type: mRNA  
A:Residues: 1-234, 'E', 236-284, 'G', 286-498, 'K', 500-542, 'K', 544-642, 'S', 644-699 <TAK>  
A:Cross-references: DDBJ:D17523; NID:94339712; PIDN:BA04477.1; PID:94339713  
A:Experimental source: liver  
C:Comment: This is a serum bactericidal factor that activates complement C4 and C2 compo  
C:Genetics:  
A:Gene: GDB:MASP1; GDB:CRAP1; CRAP1; PRSS5; MASP  
A:Cross-references: GDB:361104; GDB:330954; OMIM:600521  
A:Map position: 3q27-3q28  
C:Superfamily: complement-activating serine proteases C1r/C1s/MASP; C1r/C1s repeat homol  
C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; glycoprotein; hydro  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-448,449-699/Product: Ra-reactive factor #status predicted <MAT>  
F:19-135/Domain: C1r/C1s repeat homology <C1R1>  
F:143-181/Domain: EGF homology <EGF>  
F:185-294/Domain: C1r/C1s repeat homology <C1R2>  
F:301-362/Domain: complement factor H repeat homology <FH1>  
F:367-432/Domain: complement factor H repeat homology <FH2>  
F:449-691/Domain: trypsin homology <TRY>  
F:498-178,407/Binding site: carbohydrate (Asn) (consalent) #status predicted  
F:73-91,143-157,153-166,168-181,185-212,242-260,301-349,329-362,367-414,397-432,436-572,  
F:159/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
F:448-449/Cleavage site: Arg-Ile (autolytic) #status predicted  
F:490,552,446/Active site: His, Asp, Ser #status predicted

Query Match 14.9%; Score 565; DB 1; Length 699;  
Best Local Similarity 24.9%; Pred. No. 9.1e-30;  
Matches 190; Conservative 112; Mismatches 272; Indels 188; Gaps 31;

QY 31 SKTVQRTQDNSCSGFLHARGVELMRFTTGPDPSPYPAHARQWALRGADSVLSLFRS 90  
DB 15 SKASAHVVELANNMG-----QIOSPGVPDS-YSDSEVYNTIVPQGRFKLYPMH 64  
QY 91 FDLASCDERSDLVYNTLS-EMERHALVOLCG-----TTPSY-NLTFH 134  
DB 65 FNLSSYLCEYDYKV-----ETEDQVLAFTCGHETDTQTPQGEVVLSPGSPMSITFR 119  
QY 135 GSONVLLITLITNTERRHGFEPATFPOLP----- 163  
DB 120 SD-----FNSKEFTFPDAHYMAVDVDECKERDEBELSCDHYCHNYIGYYCSR 170  
QY 164 -----RMSGCGRLRKAQGTENSPPYPHGYPNIDCTNIEVNNQHVAFKPF 212  
DB 171 GYLITNTRCKRVCASNLTQRTGVTSDPFPRIYKSECLYTLIEBGFVNNLOPED 230  
QY 213 FYLLPFRACGTCPKYVEIN-GEK---YCGERSQPVVNSNKITVRHFSDDSYDTG 267  
DB 231 IFDIDQHPREV-PCPYDYIKIKVGPVKVLPFCGEKAPRISTQSHSVILFLPHSDMSAENRG 289  
QY 268 FLAEYLSYDSDPDP-----GQFTKRTGRCIRKELRCQGMADCTHSDDELNCSDAQH 320  
DB 290 WRLSYRA--AGNECPFLQPPVHGKIEPSQAKYFEK-----DQVLVSCDTGY 333  
QY 321 QF-----TKNKKFCPLFW-----VCDVNDGDNDEGSCGCPAQ-----TFRC 361  
DB 334 KVLKQNVEMTFTQELCKDGTWNSKIPKTCIVD-----CRAGELEHGLITSTR 383  
QY 362 NGKCLSKSQCGKNGKDDCGDGSDEASCPKXNVVYTCRHYRCLNGCLSLKGNPEC 415  
DB 384 NNLITTYKSEI---KYSQCE-----PYVGLNNTGTIYCSAQGVMMNRV-LGRSLPTC 432

QY 416 DGEDCSDGSDKDCDGLSFTRO--ARVVGTDADGEMPMQVSLHALGQGHICGASL 473  
D 433 LPV-----CGLPFRSKLMARIENGPRACKGTPWTLAMLSHNGCFGCGSL 479  
QY 474 ISPMWLSAAHCY---ID--DRGRYSQ---PTQWTFPLGLHDOSQSAFQVERRLKR 525  
D 480 LGSWMITAAHCHLQSDPDPGLRDSLDLSPDFKILLGK--WRLKSDENEGH/GVYKHT 538  
QY 526 ISHFFDFDFDYDIALLELEKPAEYSSMWPRICLPASHVFPAGKAIWYTGWGHQYGG 585  
D 539 TLHPQYPTNFENDVALVELLESFVLAFWPFLCPGPPQ--QEGANVYSSGNG-KQPLQ 595  
QY 586 TGAIILOKGEIRVINQTTTCENL--LPQQLTPRMVCVFLSGVDSQCGDGGPPLSSVEA 642  
D 596 RPFETLMEIRIPIYDHSTQCAVAPLKKKVTBMDICAGEKRGKGDACAGDSGGMVTLNR 655  
QY 643 D-GRIFQAGVSWDGCAGRNKPGVYTRLPFRWIKENTGV 683  
D 656 ERGQWYLVGTYSWGDGCKDRYGVSVYIHNNKDMIRVIGV 697

## RESULT 8

KOMPL  
plasma kallikrein (EC 3.4.21.34) precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 09-Jul-2004  
C:Accession: A36557  
R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaudien, G.; Brachpapa, L.; Rochemont  
DNA Cell Biol. 9, 737-748, 1990  
A:Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparisc  
A:Reference number: A36557; MUID:91090844; PMID:2264928  
A:Accession: A36557  
A:Molecule type: mRNA  
A:Residues: 1-638 <SEI>  
A:Cross-references: UNIPROT:P26262; GB:M5858; NID:g200358; PID:AAA6393.1; PID:g200358  
A:Note: part of this sequence, including the amino ends of both the heavy and light chain  
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w  
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11  
are linked by one or more disulfide bonds.  
C:Superfamily: coagulation factor XI: trypsin homology  
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla  
F:1-19/Domin: signal sequence #status predicted <SIG>  
F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>  
F:110-199/Domin: apple repeat <AP1>  
F:200-289/Domin: apple repeat <AP2>  
F:291-380/Domin: apple repeat <AP3>  
F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>  
F:391-621/Domin: trypsin homology <TRY>  
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322  
F:127,215,308,396,494/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 14.4%; Score 545.5; DB 1; Length 638;  
Best Local Similarity 27.5%; Pred. No. 1.6e-28;  
Matches 186; Conservative 87; Mismatches 228; Indels 175; Gaps 32;

QY 99 RGSDLVVTYNTLSMEPHALVQLGCTYPPSYNLTFHSSQWVLTLLTNRHRHGP-E 156  
D 32 RGGDLAAIYTP---DAQYCKKCTFHPRLCSF-----LAVTPEKTKRPGCFMKE 81  
QY 157 ATPFLPRMSSCG--GLRKAQGTSPYPYGHYP-----NIDCTWNIE-----V 200  
D 82 SITGTLPRIRHTGALISGHSLKQCHQISACHRDYKGLDRGSMFNISKTDNIECCQKLC 141  
QY 201 PNNQHVAVRKF---FYLLPPRPAC-----GT----- 224  
D 142 TNNFHCQF-FYVATSAFYRPRYKCKLKHASAGSTPISADNLVSGFSIKSCALSEIG 200  
QY 225 CPKD-----YVETNGSKYGEBS--QFVTSNKNITVRFHSDQSYTNG-----PLAE 271  
D 201 CPMDIFQHSAPADNLVSCVITTPDAFVCRITICTFHPNCLFTFTYNEWTESQNRVCFLKT 260

QY 272 YLSYSSDPCCPG-----FTCRTRG---CIRKEIRCDGMADCTDHSDELNCSCDAGH 320  
D 261 SKSGRSPPLIPGMALISGVSLTLCRKRPRPCHSKI-----YGVDFEGBELVTVFOGA 315  
QY 321 ---QFTCKNFKCPLRWVCDV--NDCGDNSEDEGCGCPAOTFFCSNGKLSKQCCNGKD 376  
D 316 DVCQETC-TTITRCQFFIYSLRPDC---KEGCKC--SLSLST----- 353  
QY 377 DCGDGSDEACPKAVNYTCCKHTYRCANG-----LCSKGNPECDGKEDCSDGSDKDC 430  
D 354 ---DGSP-----TRITYGMQSSGYSILRLCKLVSDCTTKIN----- 388  
QY 431 DCGLRSEFTQARVVGTDADGEMPMQVSLHA--LQGHICGASLISPMWLSAAHCYTD 488  
D 389 -----ARIVGFTNLSLGEPMQVSLQVLTAVQTHLCGSSILGRQVTLRAHCF-- 436  
QY 489 DRGFRISDPQWTAFLGLHDQO--RSAPQVQRLKRIISHFFPNDFTDYDIALLELE 546  
D 437 -DIPY--PDVWRIVYGIILSLSEITKETP---SSRIKELIHQEVKVBEGNDIALIKQ 490  
QY 547 KPAEYSSMWPRICLPASHVFPAGKAIWYTGWGHQYGTGALILQKGEIRVINQTTTCEN 606  
D 491 TLPLATYRFPQKPLCPBSKADNTITYNCWVTGMQYTKEGSETONILQKATTPVPNBECK 550  
QY 607 LRPQO-ITPRMVCVFLSGVDSQCGDGGPPLSSVADGRIFGAGVSWDGCAGRNKPG 665  
D 551 KYRDVYINKQMICAGYKGGTDACKDGGPPL-VCKHSGRWQVLGTVSWEGCGGRKDQPG 609  
QY 666 VYTRLPFRWIKENT 681  
D 610 VYTKVSEYMDMILKKT 625

## RESULT 9

KOMPL  
plasma kallikrein (EC 3.4.21.34) precursor - rat  
N:Alternate names: Fletcher factor; kininogenin; serum kallikrein  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 09-Jul-2004  
C:Accession: A39180; MUID:91129236; PMID:1993380  
R:Beaudien, G.; Rosincki-Chupin, I.; Mattei, M.G.; Mdkay, M.; Chretien, M.; Seidah, N.G.  
Biochemistry 30, 1628-1635, 1991  
A:Title: Gene structure and chromosomal localization of plasma kallikrein.  
A:Reference number: A39180; MUID:91129236; PMID:1993380  
A:Accession: A39180  
A:Molecule type: DNA  
A:Residues: 1-638 <BEA>  
A:Cross-references: UNIPROT:P14272; GB:J05315  
A:Note: the authors translated the codon GAG for residue 81 as Gln  
R:Seidah, N.G.; Ladenheim, R.; Mdkay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazure  
DNA 8, 563-574, 1989  
A:Title: The cDNA structure of rat plasma kallikrein.  
A:Reference number: A33320; MUID:90091743; PMID:2598771  
A:Accession: A33320  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-638 <SEI>  
A:Cross-references: GB:M30282; NID:g205010; PID:AAA41463.1; PID:g205011  
A:Note: part of this sequence, including the amino ends of both the heavy and light chain  
R:Pequin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.  
Biochim. Biophys. Acta 999, 103-110, 1999  
A:Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development c  
A:Reference number: S06851; MUID:90089457; PMID:2597701  
A:Accession: S06851  
A:Molecule type: protein  
A:Residues: 20-45;391-413 <PAO>  
R:Seidah, N.G.; Ladenheim, R.; Mdkay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazure  
DNA Cell Biol. 8, 563-574, 1989  
A:Title: The cDNA structure of rat plasma kallikrein.  
A:Reference number: I53041  
A:Accession: I53041  
A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 1-638 <RES>  
A:Cross-references: GB:M58590; NID:g206721; PIDN:AAA42069.1; PID:g206722  
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex with factor XIIa, which cleaves the molecule into a 11 and a 12 kDa fragment.  
C:Genetics:  
A:Gene: PK  
C:Superfamily: coagulation factor XI; trypsin homology  
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla  
F:1-19/Domin: signal sequence #status predicted <SIG>  
F:20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>  
F:20-109/Domin: apple repeat <AP1>  
F:110-199/Domin: apple repeat <AP2>  
F:200-289/Domin: apple repeat <AP3>  
F:291-380/Domin: apple repeat <AP4>  
F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>  
F:391-631/Domin: trypsin homology <TRY>  
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-356/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:356/Binding site: carbohydrate (asn) (covalent) #status predicted  
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 13.9%; Score 526.5; DB 1; Length 638;  
Best Local Similarity 27.4%; Pred. No. 2.9e-27;  
Matches 184; Conservative 85; Mismatches 232; Indels 171; Gaps 32;

Qy 99 RGSGLVTVVNTLSPEHPALVOLCGTYPPSYNLTFHSSQVLLTLITNERRHPGF-E 156  
Db 32 RGGDLAIY--TDAGHC-QKMCCTPHRCCLFSF-----LAVSPKTEKRGCEPKE 81  
Qy 157 ATFFQLRMSGCG--GRKRAQGTSPYPRGHP-----NIDCTNIE-----V 200  
Db 82 SITGTLRIHRTGASISGLKQCGHQSACHQDIYEGLDKMGSNFNISKTDSIEBCQKLC 141  
Qy 201 PNNQHV--VRFFFYLLERBRAC-----GT-----C 225  
Db 142 TNNHCGFTYATATAPRPRYKSCGLKRSSGTPTSIKYDNLVSGSLKSCALSETGC 201  
Qy 226 PKD-----YVEINGEKYCGERS--QFVTSNSNKITVRPHSDQSYDTGFLAEYLSYD 276  
Db 202 PMDIFQHPADLNVQVTPDAFVCRVCTFPHRCCLFPTFTYBEMTESQRNVCFLKTS 261  
Qy 277 SSD-PCF-----GQTCRTGR--CIKELKRDGADCTHSDDELNCCDAGH- 320  
Db 262 KSGRPSPIIGENAVSGVSLFTCKRAREPCHFKI-----YSGVAFEGEELNATFVQCAD 316  
Qy 321 --OFTCKNKKFCPLFWVCDVNDGDNDSBOGSCSPAOTFRCSNGKCLSKSQCGCKDCC 378  
Db 317 ACQETCTKTRICQFTYSLPQDC--KAECCK--SLRLST----- 353  
Qy 379 GDGSDASCPKVNVTCTKATYRCLNGLCLSKNPECDGKDCSDGSDKDC-----DC 432  
Db 354 -DGRP-----TRITY-----BAQG-----SSGVSRLCKVVSDDC 383  
Qy 433 GLRFTQARVVGTDADGEMPMQVSLHA--LGGGHCASLSLSPMLVSAACIYIDR 490  
Db 384 TTKI---NARIVGGTNSIGEMPMQVSLVQVTLVSONHCGSIIIGRWILTAACF--D 437  
Qy 491 GFRSDPTQWTAFLGHDQSO--RSAPGVGRRLKRIISHPFDFTFDYIALLEKRP 548  
Db 438 GIPY--PDWRIYGGILNLSITNKP--FSSITELTHOKYKMSBSGSIYALIKLQTP 492  
Qy 549 AEVSMVRPCLCPDASHVFPAGKAIWYMGHTQYGTGALILQKEIRVINQTTCEML 608  
Db 493 LNYTEFOKRICLPKADTNITVTNCTWYGMGYTKERGETQNILOKATIPLVNBECCQKY 552  
Qy 609 PQQ-ITPRMVCVGLSGVSDSCGDSGAPLSSVEADGAIPOAGVSWKDCGAKRKPVY 667  
Db 553 RQVYITKMTICAGYKEGIDACKDGSGLP-VCKHSGMQLVGIITSWGECARKEQPGVY 611  
Qy 668 TRLPLFRDWKE 679

Db 612 TKVAEYIDWILE 623

RESULT 10  
151468  
complement C3b/C4b inactivator (BC 3.4.21.-) precursor - African clawed frog  
N:Alternate names: C3b/C4b inactivator factor I  
C:Species: Xenopus laevis (African clawed frog)  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: 151601; S15468  
R:Kunath-Muglia, L.M.; Chang, G.H.; Sim, R.B.; Day, A.J.; Ezekowitz, R.A.  
Mol. Immunol. 30, 1249-1256, 1993  
A:Title: Characterization of Xenopus laevis complement factor I structure--conservation  
A:Reference number: 151601; PMID:94019415; PMID:7692240  
A:Accession: 151601  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-613 <KUN>  
A:Cross-references: UNIPROT:Q03711; EMBL:X59958; NID:g64595; PIDN:CAA42582.1; PID:g64596  
C:Superfamily: human complement factor I; LDL receptor ligand-binding repeat homology; tr  
C:Keywords: hydrolase; serine proteinase  
F:220-254/Domin: LDL receptor ligand-binding repeat homology <LDL1>  
F:257-290/Domin: LDL receptor ligand-binding repeat homology <LDL2>  
F:370-599/Domin: trypsin homology <TRY>

Query Match 13.4%; Score 506; DB 2; Length 613;  
Best Local Similarity 26.2%; Pred. No. 6.2e-26;  
Matches 153; Conservative 81; Mismatches 181; Indels 168; Gaps 22;

Qy 225 CPKD---YVEINGEK---YCGERSQFVTSNSNKITVRPHSDQSYDTGFLAEYLSYDS 277  
Db 65 CPKNATTEVCTDGRKQSYQQLKS--VECSNPLNSKYRRSSBAPCETFTLQ-----N 117  
Qy 278 SDP-----CEQFTCRTRGKTRKELCDGADCT----- 306  
Db 118 GEPKGIILKYVLPTEBOELPLCGKQWNSRANVVCROLGSTGADASDPSVSLYTEKP 177  
Qy 307 -DHDELNC-----SC-----DAGH-OFTCKNKKFCPLFW 334  
Db 178 PEHCTIQTGGLNSLAELCALRKLPMODNOVAKYCTTENKDCGFGFTSNGKCI9BL 237  
Qy 335 VCDVNDGDNDSBOGSCSPAOTFRCSNGKCLSKSQCGCKDCCGSDGSDASG----- 387  
Db 238 ACDSKNDGSLSDBLCKSKGNA--GFHCRSDPTCIPEQYRCNGELDLCIGEDBSNCTVBOEQ 296  
Qy 388 -----PKVNVVTCTKATYRCLNGLCLSKNPECD 416  
Db 297 KSEKQEVQKQTSKQEDLVQESKATQVEBKAKIYNYDIDARRL-----LMSLPE-- 350  
Qy 417 GKEDCSGDSBEKDCGL-----RSFTROARVVGTDADGEMPMQVSLHAIGQCHIG 470  
Db 351 -----LSCGVPPQTAALTTRKKRYVGGTNAVNKQPPMVALKD-GTAVNCG 396  
Qy 471 ASLISPMVLVSAACIYIDRGRFSDPTQWTAFLGL-----HDQSQRASGVQERRLRKI 525  
Db 397 GIYIGGCVLTAAAC-----VSNQPORVLIHMLDLRLSYKDDLSF-----YKSV 444  
Qy 526 ISHPFNDTFEDYDIALLEB-----KPAEYSMVRPCLCPDASHVFPAGKAIWYMGW 579  
Db 445 IYHETLYNPNTYENDIALLEVYNIYNNPKCMQADNNMVPACVWPSPQFKAGDTCTVSGWG 504  
Qy 580 HTQYGGTALILQKEIRVINQTTCEMLPQOITPRMVCVGLSGVSDSCGDSGAPLSS 639  
Db 505 R-EKSGMSRVFLKWKGHNTLND--NCTVYERFLDKKECARTYDGSIDACKDGSGLPVC 561  
Qy 640 VEAQGRIFQAGVSWKDCGAKRKPVYTRPLPRDWIKENTG 682  
Db 562 YDVAKVAYWGIYSWGENCGVPGVYTYTAVYFEMIAQVG 604

RESULT 11  
PLPG  
plasma (BC 3.4.21.7) precursor - pig (fragment)

N/Alternate names: plasminogen  
N/Contains: miniplasminogen  
C/Species: Sus scrofa domestica (domestic pig)  
C/Date: 07-Sep-1990 #sequence revision 01-Nov-1996 #text\_change 09-Jul-2004  
C/Accession: S03733; S03737; A25834  
R/Schaller, J.; Marti, T.; Roesele, S.J.; Kaempfer, U.; Rickli, E.E.  
Fibrinolysis 1, 91-102, 1987  
A/Title: Amino acid sequence of the heavy chain of porcine plasmin. Comparison of the ca  
A/Reference number: S03733  
A/Accession: S03733  
A/Molecule type: protein  
A/Residues: 1-560 <SCH>  
A/Cross-references: UNIPROT:P06867  
R/Brunsholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg, M.;  
Eur. J. Biochem. 114, 465-470, 1981  
A/Title: Comparison of the primary structure of the N-terminal CNBR fragments of human,  
A/Reference number: S03735; MUID:81212097; PMID:7238497  
A/Accession: S03737  
A/Molecule type: protein  
A/Residues: 1-57 <BRU>  
R/Marti, T.; Schaller, J.; Rickli, E.E.  
Eur. J. Biochem. 149, 279-285, 1985  
A/Title: Determination of the complete amino-acid sequence of porcine miniplasminogen.  
A/Reference number: A25834; MUID:85203907; PMID:3846533  
A/Accession: A25834  
A/Molecule type: protein  
A/Residues: 450-790 <MAR>  
C/Function:  
A/Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va  
ns the walls of the graafian follicle; also activates the urokinase-type plasminogen act  
A/Pathway: fibrinolysis  
C/Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology  
C/Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma; serine prote  
F.1-790/Product: plasminogen #status predicted <PRO>  
F.1-777/Domain: plasminogen-related protein precursor homology (fragment) <PLPH>  
F.1-777/Domain: activation peptide #status predicted <APT>  
F.78-560/Product: plasmin chain A #status predicted <ACH>  
F.84-167/Domain: kringle homology <KR1>  
F.166-243/Domain: kringle homology <KR2>  
F.256-333/Domain: kringle homology <KR3>  
F.358-435/Domain: kringle homology <KR4>  
F.450-790/Product: miniplasminogen #status experimental <MIN>  
F.461-540/Domain: kringle homology <KR5>  
F.561-790/Product: plasmin chain B #status experimental <BCH>  
F.561-783/Domain: trypsin homology <TRY>  
F.30-54,34-42,84-162,105-145,133-157,166-243,169-229,187-226,215-238,256-333,277-316,305  
bonds: #status predicted  
F.602,645,740/Active site: His, Asp, Ser #status predicted

Query Match 13.3%; Score 504; DB 1; Length 790;  
Best Local Similarity 25.7%; Pred. No. 1.1e-25;  
Matches 190; Conservative 80; Mismatches 228; Indels 232; Gaps 35;

74 WALRGDADSVLSLFRSPDLASCB-----RSDLVTVYNTISPM-----P 115  
Db 144 WCYTTDET-----RFYDCDIPCEDECHGSEHYEG-----IKTSGIECQSGWSOSP 195  
Qy 116 HALVOLGCTPPSYNLTFHSSQNVLLTLITNTER-----HGFEATPEQLPRMS-- 167  
Db 196 HAHGYLDSKF-PKNGLLKKNYCRN-----PDGEPRPFCFTTDNKKMEFCDIPTCTPP 247  
Qy 168 -----CGRLRKAQGTN-----SPYYPGHVPPNIDCTWNIENVNQ 204  
Db 248 PTSGPTVQCLKRGDNRYGVTVASGHTCQRMASQSPHKNRTPENPPCK-NLE----- 301  
Qy 205 HAKYRFFPFLERRACGTPKDYVEINNGKYC-----GERSQCVVTSNKSKITVR--- 257  
Db 302 -----ENYCRNPDGETAPWCYTTDSE---VRADYC 328  
Qy 258 ---HSDSYTDTGFLAEYLSYDSDPCGQFTCTGRCIRK-----LRCD 300  
Db 329 KIPSCGSSTTST---EHL-----DAVPPEQTPVAQDCYKNGSGSYGTSTTTTGKCKQ 380

Qy 301 GWADCTDHSDELNCSQDAGHOFTCKN-----KFCCKPLFWVCDSVNDGSDNSDEGCS 352  
Db 381 SWMSMTBHR-----HEKTPGNFNPAGLITMNYCR-----NPDADKSPWCYTT 421  
Qy 353 CPAQTR-CGNGKCLSKSQCCNG-----KDDCGDSDDEASCKR---VAVVT 394  
Db 422 DFRVRYWCNKKCSSETEQGVTFNPALIAQVPSVEDLSSEDMFGNGKRGKRAITVAGVP 481  
Qy 395 C-----TGHTYRCLNGLTSKGNPEC-----DQKED-----CSGSDKED 429  
Db 482 CQEWAGEPHR--HSITPEETNPAGLEKRYCNPBGDDNGPWCYTTNPQKLDYCVIPQ 539  
Qy 430 C-----DCGLRSFTQ---ARVVGTDADGEMPMQVSHALQGHICGASLISPMVLVS 481  
Db 540 CVMSFPCGKRPKEPKKCPARVVGCVSIPMSPMQSLIAYRRGHPCGGTLLISPEWLT 599  
Qy 482 AARCTIDRFRIRSDPQMTAFGLDQSRASRGVQERLKLIIHPFNDFPDVIA 541  
Db 600 AKQCL-----EKSSPSRYVILGAHE-YHAGVGQELDVSKLFEKP-----SEADIA 647  
Qy 542 LLELEKPAEYSSMWRFICLPDASHVFPAGKAIWVGHTQYGGTALILQGEIRVYIQ 601  
Db 648 LHKSSPAVITDKYIPACLPFTPVYVADRACITTMGERK-GTYGAGLKEARLPYIEN 706  
Qy 602 TTCE--NLPRQITPRMVCVGLSGVDSQGSQSGGLSVSEADRIPOAGVSWGDCA 659  
Db 707 KVCNRYEVLGKYSVPMELCAGHLAGGIDSCQSGSGLVCFEKDYILQ-GVTSWGLGCA 765  
Qy 660 QRRKPGYITLPLFRMIKE 679  
Db 766 LPRKPGYVAVSRPWTIEE 785

RESULT 12  
B61545  
plasmin (EC 3.4.21.7) precursor - sheep (fragments)  
N/Alternate names: plasminogen  
N/Contains: miniplasminogen  
C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C/Date: 28-Oct-1994 #sequence\_revision 01-Nov-1996 #text\_change 09-Jul-2004  
C/Accession: B61545; S28200  
R/Schaller, J.; Rickli, E.E.  
Enzyme 40, 63-69, 1988  
A/Title: Structural aspects of the plasminogen of various species.  
A/Reference number: A61545; MUID:89005015; PMID:3168975  
A/Accession: B61545  
A/Molecule type: protein  
A/Residues: 1-37;38-117 <SCH>  
A/Cross-references: UNIPROT:P81286  
R/Schaller, J.; Straub, C.; Kaempfer, U.; Rickli, E.E.  
Protein Seq. Data Anal. 5, 21-25, 1992  
A/Title: Complete amino acid sequence of ovine miniplasminogen.  
A/Reference number: S28200; MUID:93149995; PMID:1492092  
A/Accession: S28200  
A/Molecule type: protein  
A/Residues: 118-460 <SCH>  
C/Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology,  
C/Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; plasma; serine proteinase; zy  
F.1-37,38-117,118-460/Product: plasminogen (fragments) #status experimental <PRO>  
F.1-37/Domain: activation peptide (fragment) #status experimental <APT>  
F.38-117,118-230,231-460/Product: plasmin (fragments) #status experimental <MAT>  
F.41-118/Domain: kringle homology <KR4>  
F.118-460/Product: miniplasminogen #status experimental <MIN>  
F.132-211/Domain: kringle homology <KR5>  
F.226-460/Domain: plasmin chain B #status experimental <BCH>  
F.221-453/Domain: trypsin homology <TRY>  
F.212,315,410/Active site: His, Asp, Ser #status predicted

Query Match 13.3%; Score 503; DB 2; Length 460;  
Best Local Similarity 32.4%; Pred. No. 7.3e-26;  
Matches 154; Conservative 49; Mismatches 202; Indels 70; Gaps 18;

Qy 222 CGTCPRDYVEINCKYCGERSQFV-----TSNSKITVRPH-SDQSYTDTGFLAEYLSYD 276



```

Db      34 CEEBAOCYHNGGCGYGTSTTGTGRKCCQSMSSMIPHRHOKTPESTPNAGLTNNYCRNP 93
Qy      277 SSDPCPOQFTRICRIRKE--LRCDGADCTDHSDELNCSAHOFTCKNPF----- 328
Db      94 DADSPWCYT--TDPRVMEFCNLKKAPOAPSVENPEADCMIGIGKGYRKKATTVAGV 151
Qy      329 -CKRLFWVCSVNDGNSDQSCSCPAQTRFCNGKCLSSQOQNGKDDCGSDSDASC 387
Db      152 PCQ$-WAAQEPHNGJFTPE-----TNPRAGLKNYCRNPD-GD----- 188
Qy      388 PKVNVVTCTKATYCLNGLCISKGNPECDGKEDSCDSDE-KDCDCGLRSFTQARVVG 446
Db      189 --VNGPFCYTTNPKLTDYC---DIPQCSFDDCKPKEVKKC-----PARVVG 234
Qy      447 TDADEGMPVQVSLHALQGGHICASLISPMVLVAHCYIDDRGFRYSDPTQWTAFLGL 506
Db      235 CVATPHSMQVQSVLRRSRREHFCCGTULISPEWVLTAAHCLDSIIG-----PSFYTLIGA 289
Qy      507 HDQSRAPCVQERLRRIISHPEFNDFTPDYDIALLEKPAVSSMVRPCLPDASHV 566
Db      290 HYEMARA-SVQELIPVSRFLPEP-----SRADIALKLSSPAVITDEVIPACLPSPVYV 342
Qy      567 FPAGKAIWVGWGTQYGGTALILQKEIRVINQTTCE--NLPPQITPRMVCVGFISG 624
Db      343 VADTVCYITOMGTO--GTGVRGLKEARLPVINKYCNRYEYNGRVKSTELCAQDLAG 401
Qy      625 GVDSCGDSGGPLSSVBADEGRIFQAGVSWGSCAQRNKPQVYTRLPFRDWIKE 679
Db      402 GTDSCGDSGGPLSCFEKDKYILQ-GVTSWGLGCAAPKPGVYRVSTYVWIER 455

```

## RESULT 13

```

A47547
serine protease stubble-stubloid (EC 3.4.21.-) - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
A/Accession: A47547
R/Appel, L.F.; Proulx, M.; Abu-Shumays, R.; Hammonds, A.; Garbe, J.C.; Fribstrom, D.; Frie
Proc. Natl. Acad. Sci. U.S.A. 90, 4937-4941, 1993
A/Title: The Drosophila stubble-stubloid gene encodes an apparent transmembrane serine
A/Reference number: A47547; MUID:93281671; PMID:7685111
A/Accession: A47547
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-786 <APP>
A/Cross-references: UNIPROT:Q05319; GB:L11451; NID:G158511; PIDN:AAA28918.1; PID:G158512
C/Genetics:
A/Gene: Sb-ebd
A/Cross-references: FlyBase:FBgn0003319
C/Superfamily: serine protease stubble-stubloid; trypsin homology
C/Keywords: hydrolase; serine protease; transmembrane protein
F/61-77/Domain: transmembrane #status predicted <TM>
F/543-781/Domain: trypsin homology <TRY>

```

```

Query Match      13.2%; Score 501.5; DB 1; Length 786;
Best Local Similarity 38.8%; Pred. No. 1.6e-25;
Matches 102; Conservative 51; Mismatches 89; Indels 21; Gaps 9;

```

```

Qy      431 DCGLRSPTR-QARVVGTDADGEMPMQVSLH-----ALGGGHICGSLISPMVLVAH 484
Db      530 ECGVPTLARPEKTRIVGKSAAFGRMPQVSVRRISFSGSTHCGALINEMIAIRGH 589
Qy      485 CYIDDRGFRYSDPTQWTAFLGADQ$--QRSAPVQERRLRIISHPEFNDFTDYDIAL 542
Db      590 C-VDDLII-----SQIRLVGEYDFSHVQEQLPYTERGVAKKV-HPRYSFLTYEYDLAL 642
Qy      543 LLEKPAVSSMVRPCLPDASHVFPACKAIWVGWGTQYGGTALILQKEIRVINQ 602
Db      643 VKLEQPLFAFVHPICLPETDSL-LIGMNAVVGWGLSGGLTLPSTLQVSVPIVSN 701
Qy      603 TCENIL-----PQQTTPRM-MCVGFLSGVDSGGSDGGLSVBADGRIFQAGVSWG 657

```

```

Db      702 NCKSMFMRAGROEFIPDIFLCAGYETGQSDSCGSPQAKSQDGRFFLAGIISWGIG 761
Qy      658: CAQRNKGVTTRLPFRDWIKEN 680
Db      762 CAENLPGVCTRIKSTPMTLEH 784

```

## RESULT 14

```

KOHNP
N/Alternate names: Kininogenin; plasma prekallikrein
C/Species: Homo sapiens (man)
C/Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 09-Jul-2004
A/Accession: A00921; A37939
R/Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A/Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four te
A/Reference number: A00921; MUID:86243359; PMID:3521732
A/Accession: A00921
A/Molecule type: mRNA

```

```

A/Residues: 1-638 <CHU>
A/Cross-references: UNIPROT:P03952; GB:M13143; NID:G190262; PIDN:AAA60153.1; PID:G190263
R/McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A/Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of 1

```

```

A/Reference number: A37939; MUID:91152016; PMID:1998666
A/Accession: A37939
A/Molecule type: protein

```

```

A/Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;141-1
/260-283,'X',285;287-291,'X',293-295;314-317,'X',319-320;321-324,'X',329-333;334-339,'X',
525;538-551;562,'X',564-567;573,'X',575-576;578-583,'X',585;592-604 <MCM>
C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w

```

```

C/Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11g
are linked by one or more disulfide bonds.
C/Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal re

```

```

C/Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal re
nogen and may also play a role in the renin-angiotensin system by converting prorenin 1r
C/Genetics:

```

```

A/Gene: GDB:KXK3
A/Cross-references: GDB:127575; OMIM:229000
A/Map position: 4q35-4q35

```

```

C/Superfamily: coagulation factor XI; trypsin homology
C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inflan
F/1-19/Domain: signal sequence #status predicted <SIG>

```

```

F/20-638/Product: plasma kallikrein #status predicted <MNT>
F/20-390/Domain: plasma kallikrein heavy chain #status predicted <RCH>
F/20-109/Domain: apple repeat <AP1>

```

```

F/110-289/Domain: apple repeat <AP2>
F/200-289/Domain: apple repeat <AP3>
F/291-380/Domain: apple repeat <AP4>

```

```

F/391-638/Domain: plasma kallikrein light chain #status predicted <LCH>
F/391-621/Domain: trypsin homology <TRY>
F/21-104;47-77;51-57;111-124;131-166;141-147;201-284;227-256;231-237;292-375;322-328;383-

```

```

F/127;308;396;453;494/Binding site: carboxypeptidase (Asn) (covalent) #status experimental
F/318-347;340-345/Disulfide bonds: #status predicted
F/390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
F/434;483;578/Active site: His, Asp, Ser #status predicted

```

```

Query Match      13.2%; Score 499; DB 1; Length 638;
Best Local Similarity 26.8%; Pred. No. 1.9e-25;
Matches 189; Conservative 72; Mismatches 212; Indels 231; Gaps 36;

```

```

Qy      99 RGSGLVTVVNTLSPMERHALVOLCGTTPSYNLTFFHSSQNVLLTL-----INTERRHP 153
Db      32 RGDGVASMYT-----PNA--QYC-----OMRCTPH--PRCLLFSFLPASSINDMERFG 76
Qy      154 GF--EATFPOLPR-----MSSCGRLRKAQGFNSPYYPGHYPNID----- 193
Db      77 CFLKDSVTGLPLKVRHRTGAVSGHSLQCGHQI-----SACRDIYKGVDMAGVFNNSKVS 132
Qy      194 -----CTWNIENVNQHVKVRKFKFYLLERPRACGTCPKDYVEINGEKC--GERSQ 243
Db      133 SVEBCQRCTNNIRC---QFSYATQGFHRAVEYRNCLT-----KYSFGGTPTA 178

```



```

QY 244 FVTSNSNKITVRFRHSDQSYTDGFE-----LAELYSDSSDPG---PGQ 284
D 179 IKVLISN-----VESGFSLPKPCALSEIGHMMIIFOHLASPDVAVRLTPDA 224
QY 285 FTRTGTCIRKELRCDDWADCTHSDLNSCDAGHFTCKNFKCKRLF-----WVCD 337
D 225 FVCTCT-----ICTYHP--NC-----LFTFTYTNVKIE 250
QY 338 SV-NDCCGNSDEQG---CSCPAQTF-----RCSNGKCLSKSQOQNGKDCGDSGSD 383
D 251 SQRNVCLTKTSESTPSSSTPQENTISGYSLTCKRLTPCHSKYIP--GVDFGSELN 308
QY 384 EASCPKRVNV--TCTK-----HTYRCLNGCLSKGNPECDGKEDSGSDGDCDGLR 435
D 309 VTFYKGVNVCQETCTKMRICQFFYSL-----PEDCK---EECKCFELR 350
QY 436 -----SPTROA-----RVNGGTDADEGEMPVQSLHA 462
D 351 LSHDGSFTRLAYGTQSSGSLRLCTGDNSTCTTKTSTAVGGTSSWGEMPVQSLQV 410
QY 463 --LGQGHICGASLISPMVLVSAACVYIDRGFRYSDPTQWTAFLGLHDSQ--RSAPGVQ 518
D 411 KLTARHLCCGSLIGHQMWLTAAHCF---DGLPLQD--VWRIYSGILNLSDTKDTPEFSQ 465
QY 519 ERRLKRIISHPFENDFFDYDIALLELEKPAEYSSWVRPICLPDASHVFPAGAKIWTGW 578
D 466 ---IKEIHHQNVKSEGNHDIALIKQAPLNTYEFQKICLPBKGDPTSTYTNCAWTGW 522
QY 579 GHYQYGTGALLQKGEIRVYNQTCENLLPQ--QITPRMVCVGLSGVSGCOSDGGPL 637
D 523 GFSKEKEKQIOTNLQKNIPLVTNECCQRYQDYITRQWVACAGKXEGKQACKDSGGL 582
QY 638 SSVADGRIFQAGVSWGDCQAKRNGVYTRLPPLFRDMIKENT 681
D 583 -VCHNGMWRVLVGTITSWGBCARREGQVYTKVAEYMDILEKT 625

```

## RESULT 15

```

A29154
complement factor I (EC 3.4.21.45) precursor - human
N/Alternate names: C3b/C4b inactivator
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C/Accession: A29154; A28434; S66420
R/Cateral1, C.F.; Lyons, A.; Sim, R.B.; Day, A.J.; Harris, T.J.R.
Biochem. J. 242, 849-856, 1987
A/Title: Characterization of the primary amino acid sequence of human complement control
A/Reference number: A29154; MUID:87241401; PMID:2954545
A/Accession: A29154
A/Molecule type: mRNA
A/Residues: 1-583 <CAT>
A/Cross-references: UNIPROT:P05156; GB:Y00318
R/Goldberger, G.; Bruns, G.A.P.; Ritz, M.; Edge, M.D.; Kwiatkowski, D.J.
J. Biol. Chem. 262, 10065-10071, 1987
A/Title: Human complement factor I: analysis of cDNA-derived primary structure and assign
A/Reference number: A28434; MUID:87280021; PMID:2956252
A/Accession: A28434
A/Molecule type: mRNA
A/Residues: 1-557,'F',559-583 <GOL>
A/Cross-references: GB:302770; NID:g182606; PIND:AAA52455.1; PID:g182607
R/Ullman, C.G.; Harris, P.I.; Sultn, K.F.; Sim, R.B.; Emery, V.C.; Perkins, S.J.
FEBS Lett. 371, 199-203, 1995
A/Title: beta-Sheet secondary structure of an LDL receptor domain from complement factor
A/Reference number: S66420; MUID:95402210; PMID:7672128
A/Accession: S66420
A/Molecule type: protein
A/Residues: 258-269 <ULH>
C/Genetics:
A/Gene: GDB:IF
A/Cross-references: GDB:120077; OMIM:217030
A/Map position: 4q24-q25
C/Superfamily: human complement factor I; LDL receptor ligand-binding repeat homology; h
C/Keywords: glycoprotein; hydrolase; serine proteinase

```

```

F,1-18/Domain: signal sequence #status predicted <SIG>
F,19-335/Product: complement factor I heavy chain #status predicted <CFH>
F,221-255/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F,259-293/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F,340-583/Product: complement factor I light chain #status predicted <CFL>
F,340-569/Domain: tryptan homology <TRY>
F,70,103,177,464,494,536/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match 13.0%; Score 492.5; DB 2; Length 583;
Best Local Similarity 30.0%; Pred. No. 4,6e-25;
Matches 133; Conservative 64; Mismatches 189; Indels 57; Gaps 17;

```

```

QY 264 TDTGFLAELVSTSSD---PCPGQTCRGRIRKELRCDDWADCTD---HSDLNSC 316
D 165 TORRFLSDLSINSTCLVHCGLET-SLAECTFKRRRMGQDPADVVCYQKADSPW 223
QY 317 DADHQTCKKPKPEPLFWCDSDVNDGSDNSDEQGC--SCPAQTRGSGNGKLSQOQNGK 375
D 224 D--DFQCVNGKTI SQMKACDGINCGDSDELCKACQKGFHCKSGVCIPOYOQNGE 281
QY 376 DDCGDSDEASG-----PKVNVYCTKHTYRCLNGCLSKGNPECDGKEDSGSD 427
D 282 VDCITBEDVGCAGFASVAGEETELTADMDERRRIKSLPK----- 324
QY 428 KDCDCGL--RSFTRQARVYCGTDADGEMPVQVSLHALGCGHICGASLISPMVLVSAAC 485
D 325 --LSCVCKRMHRRRIKRIYGVGRKQAGDLPWQVALND-ASGLTCGGIYIGGCWILTAAC 381
QY 486 YIDRGFRYSDPTQWTAFLG-LHDSQSAPOQERLKRISHPFENDFFDYDIALLE 544
D 382 LRASKTHRD--IWTTVDMIHPLDKR---IVIEYDRIIHEHYNMAGTYNDIALLE 434
QY 545 LERPAEYS--SWVR--PICLPDASHVFPAGAKIWTGHTGYGTGALLQKGEIRVN 600
D 435 MKDGNKKXOCLELPRSIACVPWSPVLFQPNDTCTVSGWGR-EKDNRRVSLQGEVGLI- 492
QY 601 QTTCEMLPQQTIPR--MNCVGFSLSGVDSCQSGSGPLSVADGRIFQAGVSWKDGCA 659
D 493 -SNCSRFYGNRFYKEMECAGTYDGSIDACKDSGGLVCMANNTYVWGVVSWGECG 551
QY 660 QRKPKGVYTRLPPLFRDMIKENTG 682
D 552 KPEFPGVYTKVANYFPMISTYHVG 574

```

## RESULT 16

```

PIHU
plasma (EC 3.4.21.7) precursor [validated] - human
N/Alternate names: plasminogen precursor [nismomet]
N/Contains: angiotatin; microplasmin; plasminogen
C/Species: Homo sapiens (man)
C/Date: 24-Apr-1984 #sequence_revision 02-Dec-1994 #text_change 09-Jul-2004
C/Accession: A35229; I52242; A266738; I84609; S03735; A00929; A04625; A046
R/Petersen, T.E.; Martzen, M.R.; Ichinose, A.; Davie, E.W.
J. Biol. Chem. 265, 6104-6111, 1990
A/Title: Characterization of the gene for human plasminogen, a key proenzyme in the fibr
A/Reference number: A35229; MUID:90202879; PMID:23184848
A/Accession: A35229
A/Molecule type: DNA
A/Residues: 1810 <PBT>
A/Cross-references: UNIPROT:P00747; UNIPROT:Q9UBQ9; UNIPROT:Q9UMI2; GB:J05286; GB:M34276;
R/Malgaretti, N.; Bruno, L.; Pontoglio, M.; Candiani, G.; Meroni, G.; Ottolenghi, S.; Tar
Biochem. Biophys. Res. Commun. 173, 1013-1018, 1990
A/Title: Definition of the transcription initiation site of human plasminogen gene in liv
A/Reference number: I52242; MUID:91097523; PMID:2268308
A/Accession: I52242
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-16 <MML1>
A/Cross-references: GB:M62890; NID:g190092; PIND:AAA6454.1; PID:g553613
R/Forsgren, M.; Raden, B.; Israelsson, M.; Larsson, K.; Heden, L.O.
FEBS Lett. 213, 254-260, 1987

```

A>Title: Molecular cloning and characterization of a full-length cDNA clone for human p1  
A/Reference number: A26646; MUID:87162490; PMID:3030813  
A/Accession: A26646  
A/Molecule type: mRNA  
A/Residues: 1-471, 'D', 473-810 <FOR>  
A/Cross-references: GB:X05199; NID:G35530; PIDN:CAA28831.1; PID:G35531  
A/Experimental source: liver  
R/Malinowski, D.P.; Sadler, J.E.; Davie, E.W.  
Biochemistry 23, 4243-4250, 1984  
A>Title: Characterization of a complementary deoxyribonucleic acid coding for human and  
A/Reference number: 145961; MUID:85023311; PMID:6148961  
A/Accession: 162738  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 292-471, 'D', 473-810 <MAL2>  
A/Cross-references: GB:X02922; NID:G190112; PIDN:AAA60124.1; PID:G387031  
A/Accession: 184609  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 367-419 <MAL3>  
A/Cross-references: GB:X02921; NID:G190110; PIDN:AAA60123.1; PID:G190111  
R/Brunsholz, R.A.; Lerch, P.G.; Schaller, U.; Rickli, E.E.; Lejter, W.; Manneberg, M.;  
Eur. J. Biochem. 114, 465-470, 1981  
A>Title: Comparison of the primary structure of the N-terminal CNBr fragments of human,  
A/Reference number: S03735; MUID:81212097; PMID:7238497  
A/Accession: S03735  
A/Molecule type: Protein  
A/Residues: 20-71, 'E', 73-76 <BRU>  
R/Sottrup-Jensen, L.; Petersen, T.E.; Magnusson, S.  
submitted to the Atlas, July 1977  
A/Reference number: A00929  
A/Accession: A00929  
A/Molecule type: Protein  
A/Residues: 20-71, 'E', 73-85, 87-106, 'D', 108-360, 'E', 362-810 <SOT>  
R/Wiman, B.  
Eur. J. Biochem. 76, 129-137, 1977  
A>Title: Primary structure of the B-chain of human plasmin.  
A/Reference number: A04627; MUID:77225245; PMID:142009  
A/Accession: A04627  
A/Molecule type: Protein  
A/Residues: 581-810 <WIL>  
R/Wiman, B.; Wallen, P.  
Eur. J. Biochem. 50, 489-494, 1975  
A>Title: Structural relationship between "glutamic acid" and "lysine" forms of human pla  
A/Reference number: A04625; MUID:75093329; PMID:122932  
A/Accession: A04625  
A/Molecule type: Protein  
A/Residues: 20-50, 'Q', 51-71, 'E', 73-85, 87-100 <WIL2>  
R/Wiman, B.; Wallen, P.  
Eur. J. Biochem. 58, 539-547, 1975  
A>Title: Amino-acid sequence of the cyanogen-bromide fragment from human plasminogen cha  
A/Reference number: A04626; MUID:76043692; PMID:126863  
A/Accession: A04626  
A/Molecule type: Protein  
A/Residues: 483-507, 'E', 509-604 <WIL3>  
R/Robbins, K.C.; Bernabe, P.; Arzadon, L.; Summaria, L.  
J. Biol. Chem. 248, 1631-1633, 1973  
A>Title: The primary structure of human plasminogen. II. The histidine loop of human pla  
A/Reference number: A92125; MUID:73149248; PMID:4694729  
A/Contents: annotation; active site  
R/Groskopf, W.R.; Summaria, L.; Robbins, K.C.  
J. Biol. Chem. 244, 3590-3597, 1969  
A>Title: Studies on the active center of human plasmin.  
A/Reference number: A92048; MUID:69234739; PMID:4240117  
A/Contents: annotation; active site  
R/Trexler, M.; Vail, Z.; Patchy, L.  
J. Biol. Chem. 257, 7401-7406, 1982  
A>Title: Structure of the omega-aminocarboxylic acid-binding sites of human plasminogen.  
A/Reference number: A92382; MUID:82213905; PMID:6919539  
A/Contents: annotation; omega-aminocarboxylic acid binding sites  
R/Vail, Z.; Patchy, L.  
J. Biol. Chem. 259, 13690-13694, 1984

A>Title: The fibrin-binding site of human plasminogen. Arginines 32 and 34 are essential  
A/Reference number: A92458; MUID:85054794; PMID:6094526  
A/Contents: annotation; fibrin binding site; omega-aminocarboxylic acid binding site  
R/Geo, Y.; Ji, R.W.; Davidson, D.; Schaller, U.; Marti, D.; Soemmel, S.; McCance, S.G.;  
J. Biol. Chem. 271, 29461-29467, 1996  
A>Title: Kringle domains of human angiotensin. Characterization of the anti-proliferative  
A/Reference number: A58811; MUID:97067211; PMID:8910613  
A/Contents: annotation  
R/Lijnen, H.R.; Ugeux, F.; Bini, A.; Collen, D.  
Biochemistry 37, 4699-4702, 1998  
A>Title: Generation of an angiotensin-like fragment from plasminogen by streptolysin-1 (SM  
A/Reference number: A58812; MUID:9548733; PMID:9548733  
A/Contents: annotation  
R/Tullinsky, A.; Mulichak, A.M.  
submitted to the Brookhaven Protein Data Bank, July 1991  
A/Reference number: A51341; PDB:1PK4  
A/Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 376-454  
R/Tullinsky, A.; Wu, T.P.  
submitted to the Brookhaven Protein Data Bank, July 1991  
A/Reference number: A51488; PDB:2PK4  
A/Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 375-454  
R/Wu, T.P.; Tullinsky, A.  
submitted to the Brookhaven Protein Data Bank, August 1993  
A/Reference number: A51911; PDB:1PKR  
A/Contents: annotation; X-ray crystallography, 2.48 angstroms, residues 102-181  
R/Padmanabhan, K.; Tullinsky, A.  
submitted to the Brookhaven Protein Data Bank, April 1994  
A/Reference number: A52408; PDB:1PKK  
A/Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 377-454  
R/Tullinsky, A.; Mathews, I.I.  
submitted to the Brookhaven Protein Data Bank, December 1995  
A/Reference number: A65244; PDB:1CEA  
A/Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 102-181  
R/Tullinsky, A.; Mathews, I.I.  
submitted to the Brookhaven Protein Data Bank, December 1995  
A/Reference number: A65245; PDB:1CEB  
A/Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 102-181  
R/Mulichak, A.M.; Tullinsky, A.; Ravichandran, K.G.  
Biochemistry 30, 10576-10586, 1991  
A>Title: Crystal and molecular structure of human plasminogen kringle 4 refined at 1.9 Å  
A/Reference number: A58819; MUID:92031502; PMID:1657148  
A/Contents: annotation  
R/Wu, T.P.; Padmanabhan, K.; Tullinsky, A.; Mulichak, A.M.  
Biochemistry 30, 10589-10594, 1991  
A>Title: The refined structure of the epsilon-aminocaproic acid complex of human plasmin  
A/Reference number: A58818; MUID:92031503; PMID:1657149  
A/Contents: annotation  
R/de Vos, A.M.; Ultsch, M.H.; Kelley, R.F.; Padmanabhan, K.; Tullinsky, A.; Westbrook, M.J.  
Biochemistry 31, 270-279, 1992  
A>Title: Crystal structure of the kringle 2 domain of tissue plasminogen activator at 2.1  
A/Reference number: A39483; MUID:92118803; PMID:1310033  
A/Contents: annotation; X-ray crystallography, 2.4 angstroms  
R/Steck, B.; Teeter, M.M.; Whitlow, M.; Yamano, A.  
submitted to the Brookhaven Protein Data Bank, June 1995  
A/Reference number: A65980; PDB:1KRN  
A/Contents: annotation; X-ray crystallography, 1.67 angstroms, residues 376-454  
R/Rejzante, M.; Llinas, M.  
submitted to the Brookhaven Protein Data Bank, August 1996  
A/Reference number: A65803; PDB:1HPJ  
A/Contents: annotation; conformation by (1)H-NMR, residues 103-181  
R/Rejzante, M.R.; Llinas, M.  
submitted to the Brookhaven Protein Data Bank, August 1996  
A/Reference number: A65804; PDB:1HPK  
A/Contents: annotation; conformation by (1)H-NMR, residues 103-181  
R/Rejzante, M.R.; Llinas, M.  
submitted to the Brookhaven Protein Data Bank, August 1996  
A/Reference number: A43645; MUID:94237157; PMID:8181475  
A/Contents: annotation; conformation by (1)H-NMR, residues 96-184  
R/Rejzante, M.R.; Llinas, M.  
submitted to the Brookhaven Protein Data Bank, August 1996  
A/Reference number: A58817; MUID:94237158; PMID:8181476

A;Contents: annotation, conformation by (1)H-NMR  
 C;Comment: Plasminogen is synthesized by the kidney and is present in plasma and many of  
 C;Comment: Plasminogen is converted to plasmin by plasminogen activators (see PIR:UKHU a  
 d PIR:FGHUBB).  
 C;Comment: Plasmin is inactivated by alpha-2-antiplasmin (see PIR:IRHUX2) immediately af  
 ter-580, resulting in two chains connected by two disulfide bonds. Without the inhibitor,  
 C;Comment: Microplasmin is formed by autolytic cleavage of plasmin under artificial cond  
 C;Comment: Stromelysin 1 (see PIR:KCHUS1) acts on plasminogen to produce angiotatin. Th  
 ling solid tumors.  
 C;Genetics:  
 A;Gene: GDB:PLG  
 A;Cross-reference: GDB:119498; OMIM:173350  
 A;Map position: 6q26-6q27  
 A;Intons: 17/1; 98/1; 136/2; 183/1; 223/2; 263/1; 317/2; 366/1; 419/2; 480/1; 529  
 C;Function:  
 A;Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va  
 ries the walls of the graafian follicle; also activates the urokinase-type plasminogen act  
 A;Pathway: fibrinolysis  
 C;Superfamily: plasmin; kringe homology; plasminogen-related protein precursor homology  
 C;Keywords: angiogenesis inhibitor; blood; duplication; fibrinolysis; glycoprotein; hyd  
 F;1-96/Domain: plasminogen-related protein precursor homology <PLPH>  
 F;1-19/Domain: signal sequence #status predicted <SIG>  
 F;20-810/Product: plasminogen #status experimental <PRO>  
 F;20-96/Domain: activation peptide #status experimental <APT>  
 F;79-466/Product: angiotatin #status experimental <AST>  
 F;97-580/Domain: plasmin chain A #status experimental <MAT>  
 F;103-181/Domain: kringe homology <KR1>  
 F;185-262/Domain: kringe homology <KR2>  
 F;275-352/Domain: kringe homology <KR3>  
 F;377-454/Domain: kringe homology <KR4>  
 F;481-560/Domain: kringe homology <KR5>  
 F;550-580,581-810/Product: microplasmin #status experimental <MMT>

Query Match 13.0%; Score 492.5; DB 1; Length 810;  
 Best Local Similarity 27.0%; Pred. No. 6.3e-25;  
 Matches 194; Conservative 77; Mismatches 226; Indels 221; Gaps 39;

QY 92 DLASCDER-----GSDL-VTVYNTLSRME-----PHALVOLCTYP--PSYNTLPH 134  
 D 177 DLECEECNHCSENVNDKISKTMSGLFEOAMDSQFPA---HGYTPKFPKMLAKN 232  
 QY 135 SSQNVLLITLITNTER-----HPEGATFPQLPRMS---SCGFLRAQGTFSNPPY 184  
 D 233 YCHN-----PRLRPWCFTTDPNKRWEICDPRCTPPSSGPGYQCKKGTGE--- 281  
 QY 185 PGHYPNIDCT-----NIEVPNNQHVKKRKFYLLBPRACGCTCKDYVEIN-GE 235  
 D 282 --NIRGNVAVTSGHTCOHWSAOTPHYN-----RTPENFPCKNLDE 321  
 QY 236 KYC-----GERSQFVYVTSNSNKTITVRFHSDSYTDTGFLAETLSYSSDPCP----- 282  
 D 322 NYCENPDKKAPMCHTNSQ---VRW-----EYCKIPSCDSSPVSTEQLAAPT 365  
 QY 283 -----GQ-----FTRTGRCIRKEIKRCGMADCTDHSSELNCSQAGH 320  
 D 366 APBELTFVVDVCYHGDQSYRGTSSTTTTGK-----KCSWSWMTFHR-----H 409  
 QY 321 QPTCKNKFCKPLFWVDCSVNDCCGNSDEGSCC-----PAQTFR-CSNGKC----- 365  
 D 410 QKTPENYPAAGL-----TMYTCRNPADKGPWCFTTDPSPRWKCYCNLKCSGTEASVAP 464  
 QY 366 ----LSKSOOCNGKDC--GDG-----SDEAS-----CPKVVVVT 394  
 D 465 PAVVLLPNTVPSBEDCMFGNGKGRKARATTVTGTCQDMAQDEPHRHSITPENTPRA 524  
 QY 395 CTNHTTTCAN--GLCSKKNPEC--DKEPDCSGSDEKDC-----DGLRSFTKQ---AR 442  
 D 525 GLEKNY--CRNPDG---DVGGFWCTTTPRKLYDCVDPQCAAFBDCGKQVPEKCKPCR 580  
 QY 443 VVGSTADDEBEMPOVSLHLAGCHIGASISLNNMLVSAHNCYIDRGRFYSDPTQMTA 502  
 D 581 VVGCVAHPSHWQVSLRTRFGNHFCGGLTISPWWLTAHCL--EKSPF---PSSYKV 635

QY 503 FLGIHDQSQSAPQVQERLRKRIISHPFNDFTFDYDIALLELEKAEYSSWVRP1CLPD 562  
 D 636 ILGAAH-QEVNLEPHVQIEVSRLEFLEPTRK-----DIALTKLSSPAVITDKVIYACLPS 688  
 QY 563 ASHYFPAKRAIWTGNGHGYGGTGALLIOKGEIRYINOTTG--NLLVQOITPRMVCV 620  
 D 689 PNTVVDRTRECFITGGEIQ--GTFGAGLKEALPIENKVCNRYEFLNGRVSTELCAG 747  
 QY 621 FLSGVDSGCGSDSGGLPSVEADGRIFQAGVSWMGCAQNRKPGVYTRPLPRDWIK 678  
 D 748 HLAGTDSGCGSDSGGLVCFEKKDKYILQ--GYTSMGICARPKNGYVYVSRVITWIE 804

## RESULT 17

PLMS  
 A;Contains: angiotatin; plasminogen

C;Species: Mus musculus (house mouse)

C;Date: 20-Sep-1991 #sequence\_revision 01-Nov-1996 #text\_change 09-Jul-2004

C;Accession: A38514; S48202; S48203

R;Degeen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Billiot, R.W.  
 Genomics 8, 49-61, 1990

A;Title: Characterization of the cDNA coding for mouse plasminogen and localization of tr

A;Reference number: A38514; MUID:91184812; PMID:2081600

A;Accession: A38514

A;Molecule type: mRNA

A;Residues: 1-812 <DEG>

R;Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.  
 Eur. J. Biochem. 224, 863-871, 1994

A;Title: Characterization of the murine plasma fibrinolytic system.

A;Reference number: S48202; MUID:95010076; PMID:7523320

A;Accession: S48202

A;Molecule type: protein

A;Residues: 20-25 <LIJ>

A;Accession: S48203

A;Molecule type: protein

A;Residues: 22-27 <LIJ>

C;Comment: Plasminogen is synthesized by the kidney and is present in plasma and many of

C;Comment: Plasminogen is converted into plasmin by plasminogen activators, both plasmin

mediately after dissociation from the clot. In the presence of the inhibitor, the activat

C;Comment: Stromelysin 1 (see PIR:KCMS1) acts on plasminogen to produce angiotatin. To g

eful in treating solid tumors.

C;Function:

A;Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a vai

ns the walls of the graafian follicle; also activates the urokinase-type plasminogen acti

A;Pathway: fibrinolysis

C;Superfamily: plasmin; kringe homology; plasminogen-related protein precursor homology;

C;Keywords: angiogenesis inhibitor; blood; duplication; fibrinolysis; glycoprotein; hydric

F;1-96/Domain: plasminogen-related protein precursor homology <PLPH>

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-812/Product: plasminogen #status predicted <PRO>

F;20-96/Domain: activation peptide #status predicted <APT>

F;79-466/Product: angiotatin #status predicted <AST>

F;97-581,582-812/Product: plasmin #status predicted <MAT>

F;103-181/Domain: kringe homology <KR1>

F;185-262/Domain: kringe homology <KR2>

F;275-352/Domain: kringe homology <KR3>

F;377-454/Domain: kringe homology <KR4>

F;481-560/Domain: kringe homology <KR5>

F;582-812/Domain: chain B #status predicted <BCH>

F;582-805/Domain: trypsin homology <TRY>

F;48-73,53-61,103-181,124-164,152-176,185-262,188-316,206-245,234-257,275-352,296-335,324

bonds: #status predicted

F;78-79/Cleavage site: Glu-Asn (stromelysin 1) #status predicted

F;126,308/Binding site: carbonyl (Asn) (covalent) #status predicted

F;466-467/Cleavage site: Thr-Val (stromelysin 1) #status predicted

F;581-582/Cleavage site: Arg-Val (plasminogen activator) #status experimental

F;624,667,762/Active site: His, Asp, Ser #status predicted

Query Match 13.0%; Score 491.5; DB 1; Length 812;  
Best Local Similarity 26.3%; Pred. No. 7.4e-25;  
Matches 190; Conservative 84; Mismatches 268; Indels 181; Gaps 36;

```
Oy FTTCGFPSPPT-----PAHARCOMALRGADSVLSLTFPSFPLASCDERGSGLVTVYNT 109
   |||      |||      |      |      |      |      |      |      |
Db 165 YTTD--PKKRDYCNIPCEBECMYCGSEKKEGKISKTMSGLDCQANDSQ----- 212
Oy 110 LSPMEPHALVOLCGTYP--PSYNLTFFHSQNVLLITLINTERR-----HPGFETFF 160
   |||      |||      |||      |||      |||      |||      |||      |||
Db 213 ----SPPH----HGVIAPKFPSPKNLKNNYCHN-----PDGEPRPMCFDTDFKRWBYC 257
Oy 161 QLPWSS-----CGRLRKAQGTN-----SPYEPHYPNIDCT 195
   |||      |||      |||      |||      |||      |||      |||      |||
Db 258 DIPCTTPPPSPPTYYOCLKGRGENYRTGVSVSGTKORMSBOTPRHRNRTEPNPCK 317
Oy 196 WNIIEV-----PNNOHV-----KVRFKFYL-----LEPRACGCP-----KD 228
   |||      |||      |||      |||      |||      |||      |||      |||
Db 318 -NLBNYCRNPDEGERAPWCYTTSOLRWYCEIIFSCSSASPDSDSVPEEQTPIVVQE 376
Oy 229 VYEINGEKYCERSQFVVTNSNKITYAF--HS--DOSYTDGFLAEVLSYDSDBCPG 283
   |||      |||      |||      |||      |||      |||      |||      |||
Db 377 CYGDGQGYRTGSTSTTIIGKKCGSWAAMPFRHKSTPENPFDAILEMNYCNPDGKGPW 436
Oy 284 QFTCTGTCIKELRCGDGMADCTDHSDELNCSDAGHOFTFKNFCKPLFWVCVSVDNC 342
   |||      |||      |||      |||      |||      |||      |||      |||
Db 437 CYS--TDPFSVR-----WEYC---NLKRCSFGSGSVVELPTVSOEF-SGPSDETDCM 482
Oy 343 -GDSDEQG-----CSGPAQTFRC-----SNGLCLSKSQOCNKDKDCGPGSD 383
   |||      |||      |||      |||      |||      |||      |||      |||
Db 483 YNGKDYRKGTAVTAATPCGGMAAQEPHRHSITPTQTNPADLEKONCNPD--GD-- 537
Oy 384 EASCPKVVNVCTKHTYR-----CLNGLCISKGNPECDEKEDCSGSDPEKDCGLRSFT 438
   |||      |||      |||      |||      |||      |||      |||      |||
Db 538 -----VNGPCCYTTNPKKLKYDCDIPLCAASASFEC-GKPYV---EPKKC----- 578
Oy 439 ROARVGSTDADEGEWPQVSLHALGOG-HICGASLISPNNLVSAANCYIIDRGFRYSDE 497
   |||      |||      |||      |||      |||      |||      |||      |||
Db 579 -PGRVVCGCVANPISMWPQISLRTRFTGQHCGGTLIAFEWVLTAHCL-----EKSRP 632
Oy 498 TQMTAFGLHQDSQSASGVVERLKRILSHPFNDPTFDVILLLELKPAAEYSSWVRP 557
   |||      |||      |||      |||      |||      |||      |||      |||
Db 633 EFYKVLILGAHEEYIRGL-DVOEISVAKLILRP-----NNNDIALKLISRATTDDKVP 685
Oy 558 ICLPDASHVPAKKAIVWTGWGHTOYGSTGALLILOKEIRVINQTCENT--LPQQITPR 615
   |||      |||      |||      |||      |||      |||      |||      |||
Db 686 ACLSPSPNMVADRITCIYTGGETO-GTFGAGRLEAKOPLPIENKVCNRYVLIANNRKYST 744
Oy 616 MCWVGLSGVSDSCGDSGGGPLSSVEADGRIFOAGVSWMGDCQAQRNKGYYTTLPLFRD 675
   |||      |||      |||      |||      |||      |||      |||      |||
Db 745 ELCAQLAGVSDSCGDSGGGPLVCFEKDKYILQ-GVTSWGICARPKNKPGYVAVSRFVD 803
Oy 676 WIK 678
   |||
Db 804 WIE 806
```

RESULT 18  
KPRO coagulation factor IXa (EC 3.4.21.22) precursor - bovine  
N/Alternate names: Christmas factor  
C/Species: Bos primigenius taurus (cattle)  
C/Date: 30-Nov-1988 #sequence revision 03-Aug-1984 #text\_change 09-Jul-2004  
C/Acession: A14757; B20274; I45891; A00923  
R/Katayama, K.; Ericsson, L.H.; Enfield, D.L.; Walsh, K.A.; Neurath, H.; Davie, E.W.; Thromb. Natl. Acad. Sci. U.S.A. 76, 4990-4994, 1979  
A/TITLE: Comparison of amino acid sequence of bovine coagulation factor IX (Christmas factor) with human factor IX.  
A/Reference number: A14757; PMID:80056619; PMID:291916  
A/Molecule type: protein  
A/Residues: 1-63, 'T', 65-416 <KAT>  
A/Cross-references: UNIPROT:P00741  
A/McMullen, B.A.; Fujikawa, K.; Kisiel, W.

Biochem. Biophys. Res. Commun. 115, 8-14, 1983

A>Title: The occurrence of beta-hydroxyaspartic acid in the vitamin K-dependent blood coagulation factors IX and X.

A'Reference number: A20274; MUID:83308813; PMID:6688526

A'Accession: B20274

A'Molecule type: protein

A'Residues: 59-63, 'X', 65-69 <MCM>

R'Choo, K.H.; Gould, K.G.; Rees, D.J.G.; Brownlee, G.G.

Nature 299, 178-180, 1982

A>Title: Molecular cloning of the gene for human anti-haemophilic factor IX.

A'Reference number: I45891; MUID:82272386; PMID:6287289

A'Accession: I45891

A>Status: translated from GB/EMBL/DDBJ

A'Molecule type: mRNA

A'Residues: 52-139 <CHO>

A'CROSS-references: GB|J00007; NID:g163053; PIDN:AAA0520.1; PID:g163054

J'Hase, S.; Kawabata, S.; Nishimura, H.; Takeya, T.; Miyata, T.; Iwanaga, T.

J. Biochem. 104, 867-868, 1988

A>Title: A new trisaccharide sugar chain linked to a serine residue in bovine blood coagulation factor IX.

A'Reference number: M44565; MUID:89213999; PMID:3149657

A'Contents: annotation

A>Note: structure and location of a carbohydrate covalently bound to Ser CysComment: Factor IX is activated by factor XII, which excises the activation peptide prior to its function as a plasma proteolytic enzyme.

C'Comment: The gamma-carboxyglutamic acid residues arise by posttranslational, vitamin K dependent modification of the gamma-carboxyglutamic acid (Gla) residues and, with stromelysin, form the active site.

C'Function:

A>Description: catalyzes the proteolytic activation of coagulation factor X in the presence of factor V and phosphatidylcholine.

A'Pathway: blood coagulation intrinsic pathway

C'Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology

C'Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxylglutamate; procoagulant; coagulation factor IXA light chain #status experimental <ALC>

F.1-146/Product: coagulation factor IXA light chain #status experimental <ALC>

F.1-45/Domain: Gla domain homology (fragment) <GLA>

F.1-81-124/Domain: EGF homology <EGF>

F.1-81-124/Domain: EGF homology <EG2>

F.147-181/Product: activation peptide #status experimental <APT>

F.182-409/Product: coagulation factor IXA heavy chain #status experimental <AHC>

F.182-409/Domain: trypsin homology <TRY>

F.178,15,17,20,21,26,27,30,33,36,40/Modified site: gamma-carboxyglutamic acid (Glu) #status experimental

F.158-23,51-62,56-71,73-82,88-99,95-109,111-124,133-290,207-223,337-351,362-390/Disulfide bond #status experimental

F.53/Binding site: carboxylate (Ser) (covalent) #status experimental

F.64/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental

F.158,168,173,261/Binding site: carboxylate (Asn) (covalent) #status experimental

F.1222,270,366/Active site: His, Asp, Ser #status predicted

Db 298 NIFSFGYGVYGVGKVFNNRGRSASIIQYLKVLVDRATCLRSYFSIYSHMFCAGYHEG 357  
 QY 625 GVDSGCGDSGGPLSSVADGRIFQAGVYVWMDGCAQNRKPCVYTRLLPLFRDWIKENT 681  
 Db 358 GKDSGCGDSGGP-HVTEVEGTSFLTGLIISWGEQCAKMGKIGIYTKVSRVYVWIKEXT 413

RESULT 19  
 hepsin (EC 3.4.21.-) - human  
 C:Species: Homo sapiens (man)  
 C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 16-Aug-2004  
 C:Accession: S00845  
 R:Lehtus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.  
 Biochemistry 27, 1067-1074, 1988  
 A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain  
 A:Reference number: S00845; MUID:88205431; PMID:2835076  
 A:Accession: S00845  
 A:Molecule type: mRNA  
 A:Residues: 1-417 <LEV>  
 C:Genetics:  
 A:Gene: GDB:HPN; TMPSR1; hepsin  
 A:Cross-references: GDB:135685; OMIM:142440  
 A:Map position: 19q11-19q13.2  
 C:Superfamily: trypsin homology  
 C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein  
 F:23-45/Domain: transmembrane #status predicted <TMN>  
 F:163-400/Domain: trypsin homology <TRY>  
 F:168-204,291-359,322-338,349-381/Disulfide bonds: #status predicted  
 F:203,357,353/Active site: His, Asp, Ser #status predicted

Query Match 12.9%; Score 490; DB 1; Length 417;  
 Best Local Similarity 35.5%; Pred. No. 4.8e-25;  
 Matches 126; Conservative 47; Mismatches 132; Indels 50; Gaps 18;

QY 342 CGDNDSDQ--GCSPDQTF--RCNNGKCLSKSQCNKGD--CGDSDGDEACPRVNVTC 395  
 Db 77 CSSRNSRNVAVGLSCENGFRLALTHSELVDVRTAGANGTSGPFC--VDEGRLP----- 126

QY 396 TKHYRLCNGCLSKMPECDCKEDCGDSDG--EKDC--DCGLRSTRQARVAVGTDADGE 453  
 Db 127 --HQRLELVISV-----C---DCPRGRPLAICDCGGRKLPVD--RIVGEGDISLGR 173

QY 454 WPMQVSLHALGQGHICASLISPMVLVSAACHYIDDGFRYSPTQWTAFLGLHDSQRS 513  
 Db 174 WPMQVSLRYDG--AHLGSGSLISGPMVLTANCFPE---HNRVLSRRRVAGA--VAQAS 226

QY 514 APGVQERRLKRIISH---PPF--NDFTPDYDIALLEKPAEYSSMVRPICLPDASHVF 567  
 Db 227 PHGIQ--LGVOAVVHGGYLPFRDPNSENSNDIALVHLSSPLPLEYIQPVCILPAAGAL 285

QY 568 PAGKAVWTGMGHVGYGTGLILQKEIRVINTQEE--NLDPQITPRMCMGFLSGG 625  
 Db 286 VDGKICTVTGNGNTQYGGQAGVLOEARVPLTISDVNCAGDPYGNQIKPKPKFCAGYBEGG 345

QY 626 VDSGCGDSGGPL--SSVEADGRIFQAGVYVWMDGCAQNRKPCVYTRLLPLFRDWI 677  
 Db 346 IDACGSGGPFVCEIDISRRPRRLGIGYSWGGCALAKPGYTTKVSDFREWI 400

RESULT 20  
 A57014  
 prostaasin (EC 3.4.21.-) precursor - human  
 C:Species: Homo sapiens (man)  
 C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
 C:Accession: A57014; A54866  
 R:Yu, J.X.; Chao, L.; Chao, J.  
 J. Biol. Chem. 270, 13483-13489, 1995  
 A:Title: Molecular cloning, tissue-specific expression, and cellular localization of human  
 A:Reference number: A57014; MUID:95286644; PMID:7768952  
 A:Accession: A57014  
 A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
 A:Residues: 1-343 <RES>  
 A:Cross-references: UNIPROT:Q16651; GB:U41351; NID:9862304; PIDN:AA41759.1; PID:9862305  
 A:Experimental source: prostate  
 A>Note: parts of this sequence were determined by protein sequencing  
 R:Yu, J.X.; Chao, L.; Chao, J.  
 J. Biol. Chem. 269, 18843-18848, 1994  
 A:Title: Prostaasin is a novel human serine proteinase from seminal fluid. Purification,  
 A:Reference number: A54866; MUID:94308140; PMID:8034638  
 A:Accession: A54866  
 A:Molecule type: protein  
 A:Residues: 45-64 <YVA>  
 C:Genetics:  
 A:Gene: GDB:PRSS8  
 A:Cross-references: GDB:676446; OMIM:600823  
 A:Map position: 16p11.2-16p11.2  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein  
 F:1-32/Domain: signal sequence #status predicted <SIG>  
 F:33-44,45-343/Product: prostaasin #status predicted <MAT>  
 F:33-44/Domain: prostaasin light chain #status predicted <CHL>  
 F:45-343/Domain: prostaasin heavy chain #status predicted <CHH>  
 F:45-281/Domain: trypsin homology <TRY>  
 F:323-341/Domain: transmembrane #status predicted <TMN1>  
 F:37-154,70-86,168-244,201-223,234-262/Disulfide bonds: #status predicted  
 F:85,134,238/Active site: His, Asp, Ser #status predicted  
 F:159/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 12.8%; Score 484.5; DB 1; Length 343;  
 Best Local Similarity 39.3%; Pred. No. 9.1e-25;  
 Matches 105; Conservative 42; Mismatches 95; Indels 25; Gaps 8;

QY 424 GSDEKDCDGLRSFTRQARVAVGTDADGEMPMQVSLHALGQGHICASLISPMVLVSA 483  
 Db 29 GAGGAEPGV--AQARITGSSSAVAGQMPQVISTYEGV--HVGGSILVSGQWLISA 84

QY 484 HCYTDGRFRYSPTQWTAFLGLHDSQRSAPGVQERRLKRIISHPPENDFTDYDIAL 543  
 Db 85 HCPESH-----HKEAYEVLGNH--QLDSYSEDAKSTLKDILPHPSYLOEGSQGIAL 138

QY 544 ELKPAEYSSMVRPICLPDASHVFPAKAIWYGMKHTQYGGTGLI-----LQKEIRV 599  
 Db 139 QLSRPITFSRYRIFCLPANNASFNGHCTYGMGHV--APSVSLTTPKPLDQLEVP 196

QY 600 NQTTGENTL-----PQGITPRMCMGFLSGVDSGCGDSGGPLSSVEADGRIFQAG 651  
 Db 197 SRTCNCLVNIADKPEPHVQEDWVCAGYBEGKXACGDSGGL--SCVEGLWYLTGI 255

QY 652 VSWGDCGAQRNKGVTYTRLLPLFRDWIK 678  
 Db 256 VSWGDCAGANRPVYTLASSVYASW 282

RESULT 21  
 KFHU  
 coagulation factor IXa (EC 3.4.21.22) precursor [validated] - human  
 N:Alternate names: antithemophilic factor B; Christmas factor  
 C:Species: Homo sapiens (man)  
 C>Date: 17-Dec-1982 #sequence\_revision 30-Jun-1987 #text\_change 09-Jul-2004  
 C:Accession: A00922; A37570; A30511; A32989; A22673; A21337; A37546; A30623; A60486; A20;  
 R:Yoshitake, S.; Schach, B.G.; Foster, D.C.; Davie, E.W.; Kurachi, K.  
 Biochemistry 24, 3736-3750, 1985  
 A:Title: Nucleotide sequence of the gene for human factor IX (antithemophilic factor B).  
 A:Reference number: A00922; MUID:86000558; PMID:2994716  
 A:Accession: A00922  
 A:Molecule type: DNA  
 A:Residues: 1-461 <YOS>  
 A:Cross-references: UNIPROT:P00740; GB:K02402; NID:9182612; PIDN:AAB59620.1; PID:9182613  
 R:Anson, D.S.; Choo, K.H.; Rees, D.J.G.; Giannelli, F.; Gould, K.; Huddleston, J.A.; Brov  
 EMO J. 3, 1053-1060, 1984  
 A:Title: The gene structure of human anti-haemophilic factor IX.  
 A:Reference number: A37570; MUID:84236100; PMID:6329734  
 A:Accession: A37570

A/Molecule type: DNA  
A/Residues: 1-461 <ANS>  
A/Cross-references: GB:K02048  
R:Reltama, P.H.; Bertina, R.M.; Ploos van Amstel, J.K.; Riemens, A.; Briet, E.  
Blood 72, 1074-1076, 1988  
A/Title: The putative factor IX gene promoter in hemophilia B Leyden.  
A/Reference number: A30511; MUID:88327116; PMID:3416069  
A/Accession: A30511  
A/Molecule type: DNA  
A/Residues: 8-24 <RE>  
A/Cross-references: EMBL:X55008; NID:G311288; PIDN:CA838245.2; PID:G4469253  
R:Koebel, D.D.; Botema, C.D.K.; Buerschede, J.M.; Sommer, S.S.  
Am. J. Hum. Genet. 45, 448-457, 1989  
A/Title: Functionally important regions of the factor IX gene have a low rate of polymor-  
A/Reference number: A32989; MUID:88371752; PMID:2773937  
A/Accession: A32989  
A/Status: not compared with conceptual translation  
A/Molecule type: DNA  
A/Residues: 30-92 <KOB>  
R:McGraw, R.A.; Davis, L.M.; Noyes, C.M.; Lundblad, R.L.; Roberts, H.R.; Graham, J.B.; S  
Proc. Natl. Acad. Sci. U.S.A. 82, 2847-2851, 1985  
A/Title: Evidence for a prevalent dimorphism in the activation peptide of human coagulab  
A/Reference number: A22673; MUID:85190593; PMID:3857619  
A/Accession: A22673  
A/Molecule type: mRNA  
A/Residues: 1-193, 'T', 195-461 <MCG>  
A/Cross-references: GB:M11309; NID:G180552; PIDN:AAA52023.1; PID:G180553  
A/Note: the authors translated the codon ACA for residue 29 as Tyr  
R:Jaye, M.; de la Salle, H.; Schamber, F.; Balland, A.; Kohli, V.; Finkel, A.; Tolstosh  
Nucleic Acids Res. 11, 2325-2335, 1983  
A/Title: Isolation of a human anti-haemophilic factor IX cDNA clone using a unique 52-b  
A/Reference number: A21337; MUID:83220786; PMID:6687940  
A/Accession: A21337  
A/Molecule type: mRNA  
A/Residues: 1-193, 'T', 195-461 <JAY>  
A/Cross-references: GB:J00137; NID:G182610; PIDN:AAA52763.1; PID:G182611  
R:Jagadeeswaran, P.; Lavelle, D.E.; Kaul, R.; Mohandas, T.; Warren, S.T.  
Somat. Cell Mol. Genet. 10, 465-473, 1984  
A/Title: Isolation and characterization of human factor IX cDNA: identification of Tag I  
A/Reference number: A37546; MUID:84300526; PMID:6089357  
A/Accession: A37546  
A/Molecule type: mRNA  
A/Residues: 38-193, 'T', 195-326 <JAG>  
A/Cross-references: GB:M35672  
R:Kurachi, K.; Davis, E.W.  
Proc. Natl. Acad. Sci. U.S.A. 79, 6461-6464, 1982  
A/Title: Isolation and characterization of a cDNA coding for human factor IX.  
A/Reference number: A30623; MUID:83065193; PMID:6959130  
A/Accession: A30623  
A/Molecule type: mRNA  
A/Residues: 1-12, 'S', 14-73, 'P', 75-82, 'K', 84-203, 'P', 205-216, 'G', 218-298, 'A', 299-356, 'A',  
A/Cross-references: GB:J00136; NID:G182608; PIDN:AAA98726.1; PID:G182609  
R:Tharakan, J.; Strickland, D.; Burgess, W.; Drohan, W.N.; Clark, D.B.  
Vox Sang. 58, 21-29, 1990  
A/Title: Development of an immunofluorescence process for factor IX purification.  
A/Reference number: A60486; MUID:90194857; PMID:2316207  
A/Accession: A60486  
A/Molecule type: protein  
A/Residues: 47-52, 'XX', 55-60, 'X', 62, 'XX', 65 <THA>  
R:McMullen, B.A.; Fujikawa, K.; Kistiel, W.  
Biochem. Biophys. Res. Commun. 115, 8-14, 1993  
A/Title: The occurrence of beta-hydroxyaspartic acid in the vitamin K-dependent blood co  
A/Reference number: A20274; MUID:83308813; PMID:6688526  
A/Accession: A20274  
A/Molecule type: protein  
A/Residues: 105-109, 'X', 111-115 <MCM>  
R:Balland, A.; Faure, T.; Carvallo, D.; Cordier, P.; Ulrich, P.; Fournet, B.; de la Sall  
Eur. J. Biochem. 172, 565-572, 1988  
A/Title: Characterisation of two differently processed forms of human recombinant factor  
A/Reference number: S02527; MUID:88166735; PMID:3280312  
A/Accession: S02527  
A/Molecule type: protein

A/Residues: 29-63 <BAL>  
A/Note: processed forms expressed in recombinant system  
R:Jallat, S.; Perraud, F.; Dalemans, W.; Balland, A.; Diesterle, A.; Faure, T.; Meulien, I  
EMBO J. 9, 3295-3301, 1990  
A/Title: Characterization of recombinant human Factor IX expressed in transgenic mice and  
A/Reference number: S12058; MUID:91006024; PMID:2209546  
A/Accession: S12058  
A/Molecule type: mRNA; protein  
A/Residues: 1-68 <JAL>  
A/Note: processed forms expressed in recombinant system  
R:Handford, P.A.; Baron, M.; Mayhew, M.; Willis, A.; Beesley, T.; Brownlee, G.G.; Campbell  
EMBO J. 9, 475-480, 1990  
A/Title: The first EGF-like domain from human factor IX contains a high-affinity calcium  
A/Reference number: S12377; MUID:90151623; PMID:2406129  
A/Accession: S12377  
A/Molecule type: protein  
A/Residues: 92-130 <HAN>  
A/Note: NMR detection of calcium binding by domain expressed in recombinant system  
R:de la Salle, C.; Charanfar, J.L.; Baas, M.J.; Schwartz, A.; Wiesel, M.L.; Grunbaum,  
Thromb. Haemost. 70, 370-377, 1993  
A/Title: A deletion located in the 3' non translated part of the factor IX gene responsi  
A/Reference number: I59612; MUID:94054330; PMID:8236150  
A/Accession: I59612  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 444-461 <RES>  
A/Cross-references: GB:S66752; NID:G439773; PIDN:AA828588.1; PID:G439774  
R:Stoffel, E.S.; Koebel, D.D.; Sarkar, G.; Sommer, S.S.  
Science 239, 491-494, 1988  
A/Title: Genomic amplification with transcript sequencing.  
A/Reference number: I59529; MUID:88127096; PMID:3340835  
A/Accession: I59529  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 290-359 <RE2>  
A/Cross-references: GB:M19063; NID:G182622; PIDN:AAA52456.1; PID:G182623  
R:Agarwala, K.L.; Kawabata, S.; Takao, T.; Murata, H.; Shimomishi, Y.; Nishimura, H.; Iw  
Biochemistry 33, 5167-5171, 1994  
A/Title: Activation peptide of human factor IX has oligosaccharides O-glycosidically lin  
A/Reference number: A54255; MUID:94227047; PMID:8172852  
A/Accession: A54255  
A/Molecule type: protein  
A/Residues: 'G', 204, 'X', 206-211, 212, 'D', 214, 'X', 216-221, 'D' <AGA>  
A/Note: the residues designated 'X' were determined to be threonine bound to carbohydrate  
R:Di Scipio, R.G.; Kurachi, K.; Davis, E.W.  
U. Clin. Invest. 61, 1528-1538, 1978  
A/Title: Activation of human factor IX (Christmas factor).  
A/Reference number: A18483; MUID:78194509; PMID:659613  
A/Contents: annotation; activation; active site; carbohydrate binding  
R:McGraw, R.A.; Davis, L.M.; Noyes, C.M.; Graham, J.B.; Roberts, H.R.; Stafford, D.W.  
Am. Soc. Hematol. Abstr. 64(Suppl.1), 262a, 1984  
A/Reference number: A37569  
A/Contents: annotation  
A/Note: 194-Thr was also found  
R:Mortle, T.; Isaacs, B.S.; Esmon, C.T.; Johnson, A.E.  
J. Biol. Chem. 259, 5698-5704, 1984  
A/Title: Derivatives of blood coagulation factor IX contain a high affinity Ca2+-binding  
A/Reference number: A37543; MUID:84185715; PMID:6425256  
A/Contents: annotation; calcium binding  
R:Mortle, T.; Isaacs, B.S.; Esmon, C.T.; Johnson, A.E.  
J. Biol. Chem. 260, 2583, 1985  
A/Reference number: A37544  
A/Contents: annotation; calcium binding, correction  
R:Bentley, A.K.; Rees, D.J.G.; Rizza, C.; Brownlee, G.G.  
Cell 45, 343-348, 1986  
A/Title: Defective propeptide processing of blood clotting factor IX caused by mutation c  
A/Reference number: A37545; MUID:86189947; PMID:3009023  
A/Contents: annotation; signal sequence cleavage site  
R:Suehiro, K.; Kawabata, S.; Miyata, T.; Takeya, H.; Takamatsu, J.; Ogata, K.; Kamiya,  
U. Biol. Chem. 264, 21257-21265, 1989  
A/Title: Blood clotting factor IX B(M) Nagoya: substitution of arginine 180 by tryptophan  
A/Reference number: A30622; MUID:90078229; PMID:2592373  
A/Contents: annotation; sequence of mutant B(M) Nagoya



A>Note: carboxylation, glycosylation, and cleavage sites  
R:Baron, M.; Norman, D.G.; Harey, T.S.; Hanford, P.A.; Mayhew, M.; Tse, A.G.D.; Brownle  
submitted to the Brookhaven Protein Data Bank, November 1991  
A:Reference number: A51252; PDB:1IXA  
A:Contents: annotation; conformation by (1)H-NMR, residues 92-130  
A>Note: recombinant form expressed in yeast  
C:Comment: Factor IX is activated by factor XIa, which excises the activation peptide pr  
C:Comment: The gamma-carboxyglutamic acid residues arise by posttranslational, vitamin K  
C:Comment: Calcium binds to the gamma-carboxyglutamic acid (Gla) residues and, with strc  
C:Genetics:  
A:Gene: GDB:F9  
A:Cross-references: GDB:119900; OMIM:306900  
A:Map position: Xq27.1-Xq27.2  
A:Functions: 30/1; 84/2; 93/1; 131/1; 174/1; 241/3; 280/1  
A:Function:  
A:Description: catalyzes the proteolytic activation of coagulation factor X in the prese  
C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology  
C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglutan  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-46/Domain: propeptide #status experimental <PPT>  
F:31-91/Domain: Gla domain homology <GLA>  
F:47-131/Product: coagulation factor IXa light chain #status experimental <ALC>  
F:97-128/Domain: EGF homology <EG1>  
F:134-170/Domain: EGF homology <EG2>  
F:192-226/Domain: activation peptide #status experimental <ACT>  
F:227-461/Product: coagulation factor IXa heavy chain #status experimental <AHC>  
F:227-454/Domain: trypsin homology <TRY>  
F:53-54,61,63,66,67,72,73,76,78,82,86/Modified site: gamma-carboxyglutamic acid (Glu) #  
F:64-69,97-108,102-117,119-128,134-145,141-155,157-170,178-335,252-268,382-396,407-435/D  
F:99/Binding site: carbohydrate (Ser) (covalent) #status experimental  
F:110/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental  
F:191-192/Cleavage site: Arg-Ala (coagulation factor IXa) #status experimental  
F:203,213/Binding site: carboxylate (Asn) (covalent) #status experimental  
F:205,215/Binding site: carbohydrate (Thr) (covalent) #status experimental  
F:226-227/Cleavage site: Arg-Val (coagulation factor IXa) #status experimental

Query Match 12.8%; Score 484.5; DB 1; Length 461;  
Best Local Similarity 27.6%; Pred. No. 1.2e-24;  
Matches 145; Conservative 76; Mismatches 184; Indels 121; Gaps 19;

Qy 188 YPRPIDCTNMIENVNQHVKYRFFYLPRRACGTCPPDYAEINEKTC-GERSQPV 246  
Db YLSAECTVFLDHNANKI-----LNRPKRNKSLKEFVQGNREMEKCSFEE 73  
Qy 247 TSNSNKTIVFPHSQSTDTGFLAEVLSYD--SSDPCPGQFTCRG----- 290  
Db 74 ARVFEENTER-----TFEFKQIVDGDQCESNPLNGSCKDIDNSIECMCPFEG 125  
Qy 291 -----RCIRKELRCGDWADCTDHS-D-ELNCSGDAHQFTCKNFKPFLWVCDVND 342  
Db 126 KNCLELDVTNKNKRCQPF--CKNSADNKVVCSTEGVRLAENKSCSE----- 171  
Qy 343 GDNDSDEGSCCPAOTFCGNSGKCLSKSQCKGKDCGDSDEASCPKVNVTCTKHTYRC 402  
Db 172 -----PAVFPFCGRVSVQTSKLTTRA-----EAVFPVDVYNSTE----- 206  
Qy 403 INGLCLSKNPECCGKSDGSDGDEKDCGLRSFTQARVVGSTDADEGEMPVQSLHA 462  
Db 207 -----ATTIIDNTQ-----STGSFNDFTRVGGEDAKPGQFPQVVLNG 246  
Qy 463 LGQGHICGASLISPMNLVSAACHYIDRGFRYSPTQWTAFLGLHDSQSAAPVQERRL 522  
Db 247 KVDA-FCGGSIVKEKIVTAHCV--ETGVKI-----TVVAGEHN-IEBTEHTEQGRNV 296  
Qy 523 KRITSHFPNDF--TFPYDIALLELEKPAEYSSMVRPICLPDASHV-----FPAKGIWV 575  
Db 297 IRIIPHNNAALINKYHNDIALLELDEPLVLSYVTEPICADKEVYTIPLKFGSG---YV 353  
Qy 576 TGMGHTQYGTGALILKGEIRVINOCTCENTLPOQITPRMVCVGFSGGVDSQSGSG 635  
Db 354 SGMGKRVHKGKRSALVLOTLRVPLVDRAITCLARSTFTTYNNMFCAGFHEGGSDSCGSDSG 413

Qy 636 PLSSVEADGRIFQAGVSWGCGAQRNKPQVYTRLEPLFMDIKENT 681  
Db 414 P-HVTEVEGTSFLTGILISWGEBCAMKKGKGIYTKVSRVYMIKEKT 458

RESULT 22  
146260  
plasma (EC 3.4.21.7) precursor - western European hedgehog  
C:Species: Erinaceus europaeus (western European hedgehog)  
C:Date: 21-Feb-1997 #sequence\_rev150n 21-Feb-1997 #text\_change 09-Jul-2004  
C:Accession: I46260  
R:Laun, R.M.; Boomark, N.W.; Schwartz, K.; Lindahl, G.E.; Wade, D.P.; Byrne, C.D.; Fong,  
J. Biol. Chem. 270, 24004-24009, 1995  
A:Title: The recurring evolution of lp(a): Insights from cloning of hedgehog apolipoprote  
A:Reference number: I46259; PMID:96025778; PMID:7592597  
A:Accession: I46260  
A:Status: preliminary; translated from GB/EMBL/DDJ  
A:Molecule type: mRNA  
A:Residues: 1-810 <LAW>  
A:Cross-references: UNIPROT:Q29485; EMBL:U03171; NID:G1046360; PID:G1046361  
C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology;  
C:Keywords: hydrolase; serine protease  
F:1-96/Domain: plasminogen-related protein precursor homology <PLPH>  
F:103-181/Domain: kringle homology <KR1>  
F:185-262/Domain: kringle homology <KR2>  
F:275-352/Domain: kringle homology <KR3>  
F:379-456/Domain: kringle homology <KR4>  
F:482-561/Domain: kringle homology <KR5>  
F:582-803/Domain: trypsin homology <TRY>

Query Match 12.7%; Score 483; DB 2; Length 810;  
Best Local Similarity 26.7%; Pred. No. 2.7e-24;  
Matches 170; Conservative 78; Mismatches 205; Indels 184; Gaps 31;

Qy 192 IDC-TNNIRVN-----NQHYKVRKFFYLPRRACGTCCK-----DYVEI----- 232  
Db 204 LECQPMDSQIPHGHGTFPSKFPKLNKMNCRNPDGSPRCWCTFMDNRKMEYCDIPRCT 263  
Qy 233 -----NGEKYCGERSQFVYTSNSNKTIVFPHSQS-----YTDG 267  
Db 264 TPRPSPGPTVQCMLNGEHNQGN-----VAVVYSGLTQCRWGEQSHRDRTPENYPCKN 318  
Qy 268 FLAEVLSYDSDPCPGQFTCRGTCIRKEL-----RCDGMACTDHSDE----- 311  
Db 319 IDENYCRNPDGEPAPMCF--TWSSVRWEBCKIPDCVSSASETHSDAPVYVPEQTPV 376  
Qy 312 LNSGCDAGHQF-----TCKNFKCKPLFWY-----CD-SVNDGSDSDE 348  
Db 377 QECYQNGQYRTGTSSTTTTGKQCP--WTSNRPHRHSKTPENYPPADLTNMYCRNPDGD 434  
Qy 349 QGCSGSC-----PAQTFR-CSNGKCL-----SKSQOC--NGKDDCGDG 381  
Db 435 KGPWCYTTTPSVAMERCNLKCGSTEMSATNSSPVGVSSASBESBDCLIDNKGKRGTK 494  
Qy 382 SDEAS-----CPKVNVTCTKHTYRC-----INGCLSKGNPE--- 414  
Db 495 ATTGAGTTCQAMAARPHRHSITFPETNPRAIDQENY-CRNPDGANDGMPCTTNRKLP 553  
Qy 415 --CDGKEDSGSDGDEKDCGLRSFTQARV-----YCGNDADGEMPMQVSLHALG 464  
Db 554 DYCDIPIHCVPSS--ADCG-----KRPVEPKKCPRGVGGCVAPHPSPMVVSLRRFG 603  
Qy 465 QGHICGASLISPMNLVSAACHYIDRGFRYSPTQWTAFLGLHDSQSAAPVQERRL 524  
Db 604 Q-HFCGTLISPMNVVTAHCL-----EKFSNAIYKVVLAGH-QETRLERDVQIGYVK 656  
Qy 525 IISHFPNDFPYDIALLELEKPAEYSSMVRPICLPDASHVFPAGKAIWVTGHTQYG 584  
Db 657 MFLEP-----YRADIALALKLSSPAITTDKHPACCPNSNMYVADRLCYITGEGSTK-G 709  
Qy 585 GTGATLLKGEIRVINOCT--ENLLPQGITPRMVCVGFSGGVDSQSGSGSLSSVEA 642  
Db 710 TYGAGLLKEAQPLVLENKVCNRSFLNGRSTELCAHGLAGGVDSQSGSGSLVCFEK 769



Qy 643 DGRIFQAGVSWGDCGQARNKRGVYTRLEPRDIKE 679  
Db 770 DRYILQ-GVTSWGLGCAKLTTRPGVYRVSRVSWLQD 805

## RESULT 23

PLBO

plasma (EC 3.4.21.7) precursor - bovine

N/Alternate names: plasminogen

C/Species: Bos primigenius taurus (cattle)

C/Date: 30-Sep-1997 #sequence revision 28-Apr-1995 #text\_change 09-Jul-2004

C/Accession: S45046; A25835; I45961; S03736

R/Berglund, L.; Andersen, M.D.; Petersen, T.E.

submitted to the EMBL Data Library, May 1994.

A/Description: Cloning and characterization of the bovine plasminogen cDNA.

A/Reference number: S45046

A/Accession: S45046

A/Molecule type: mRNA

A/Residues: 1-812 <BR>

A/Cross-references: UNIPROT:P06668; EMBL:X79402; NID:g494962; PIND:CA55939.1; PID:g4949

A/Experimental source: liver

A/Note: It is uncertain whether Met-1 or Met-8 is the initiator

R/Schaller, J.; Moser, P.W.; Danneberg-Muller, G.A.K.; Rosselet, S.J.; Kampfer, U.; Rick

Eur. J. Biochem. 149, 267-278, 1985

A/Title: Complete amino acid sequence of bovine plasminogen. Comparison with human plas

A/Reference number: A25835; MUID:85203906; PMID:3846532

A/Accession: A25835

A/Molecule type: protein

A/Residues: 27-334, 'D', 336-515, 'H', 517-554, 'L', 556-812 <SCH>

R/Malinowski, D.P.; Sadler, J.E.; Davie, E.W.

Biochemistry 23, 4243-4250, 1984

A/Title: Characterization of a complementary deoxyribonucleic acid coding for human and

A/Reference number: I45961; MUID:85023311; PMID:6148961

A/Accession: I45961

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 706-743, 'R', 745-812 <MAL>

A/Cross-references: GB:K02935; NID:g163551; PIND:AAA30714.1; PID:g163552

R/Brunsholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergler, W.; Manneberg, M.;

Eur. J. Biochem. 114, 465-470, 1981

A/Title: Comparison of the primary structure of the N-terminal CNBR fragments of human,

A/Reference number: S03735; MUID:81212097; PMID:7238497

A/Accession: S03735

A/Molecule type: protein

A/Residues: 27-83 <BRU>

C/Function:

A/Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va

ns the walls of the granular follicle; also activates the urokinase-type plasminogen act

A/Pathway: fibrinolysis

C/Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology

C/Keywords: duplication; fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma;

F/1-26/Domain: signal sequence #status predicted <SIG>

F/8-103/Domain: plasminogen-related protein precursor homology <PLPH>

F/27-812/Domain: plasminogen #status experimental <PRO>

F/27-103/Domain: activation peptide #status experimental <APT>

F/104-983,584-812/Domain: plasmin #status experimental <MPT>

F/110-583/Domain: plasmin chain A #status experimental <ACH>

F/110-583/Domain: plasmin chain A #status experimental <ACH>

F/192-269/Domain: kringle homology <KR2>

F/282-359/Domain: kringle homology <KR3>

F/384-461/Domain: kringle homology <KR4>

F/485-564/Domain: kringle homology <KR5>

Matches 151; Conservative 73; Mismatches 220; Indels 109; Gaps 24;

Qy 181 SPYRPHYPPNIDCTNIEV-----PNNQ-----HYVRFKFFL----- 215

Db 310 TPHKMRTPEPCK-NLEENYCRNPKGKAPWCYTTSNRYRMYCTIPGCESSPLSTER 368

Qy 216 -----LEPRACGTPKDYVEINKEKYGERSQFV--TSNSKTIYVFF--SDQSYDTT 266

Db 369 MDVAVPEPQ--PVPQCHNGNGSYGTSSTITTKGCKQSWGSMTHRLKTPENYRNA 426

Qy 267 GLFAEVLSDSPDPCQFT-----CRTRCIR--KELRCGMACTDHSDELNCS 315

Db 427 GLTMNCRNPDADKSPWCYTTRVRFECNLKCKSETPEQVPAAPAPQVPEMPPEADCK 486

Qy 316 CDAGHOTCKNKR-----CKPLFWGDSVNDGDSNDEGCGSPATPFCNSCKLCKS 369

Db 487 IGTGKSYRGKAKATTAGVPCOE--WAAQEPHQHSIFPE-----TNPQGLER 532

Qy 370 QQNGKDCDGDSDASCPKRVNVTCTKHYRCINGLCLSKNPEDCGKEDSGDSDE-X 428

Db 533 NYCNRP--GD-----VNGPWCTYNNPKRPFDC--DVPQCESSPDCGKPKYEPK 578

Qy 429 DDCGLRSFTRQARVGTADDEGEMFQVSLHALGGHICGASLSPNNLVSAARCYID 488

Db 579 KC-----SGRIVGCVSKPSHWPQVSLRRSR--HFCGGTLISPKWVLTAAHCLDN 628

Qy 489 DGRGRYDPTQWRAFLGLHQSGRSAPGVORRLKRIISHPFNDFPDVIALLEKRP 548

Db 629 ILALSF-----YKVLGAHNEKARE-OSVGEIPVSRLEP--SQADILALKLSRP 676

Qy 549 AEYSMWPICLPDASHVPAGKAIWYTGNGHYQYGGTGAIILOKGEIRYINQTT--EN 606

Db 677 AITKEVTPACLPENNVAAARTECYITGGERQ--GRFGGLKKEALPIPIENKVCNRNE 755

Qy 607 LLEQITPRMVCVGLSGVDSGSGSPVSEADGRIFQAGVSWGDCGQARNKRGV 666

Db 736 YLDGRVKTPELCAHGLIGTDSGSGDGLVCEKDKYILQ-GVTSWGLGCAKLTTRPGV 794

Qy 667 YTRLEPRDIKE 679

Db 795 YVRVSPVPIER 807

## RESULT 24

plasma (EC 3.4.21.7) precursor - thesus macaque

C/Species: Macaca mulatta (thesus macaque)

C/Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 09-Jul-2004

C/Accession: B32869; B30848

R/Tomlinson, J.E.; McLean, J.W.; Lamn, R.M.

J. Biol. Chem. 264, 5957-5965, 1989

A/Title: Rhesus monkey apolipoprotein(a). Sequence, evolution, and sites of synthesis.

A/Reference number: A32869; MUID:89174660; PMID:2925643

A/Accession: B32869

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-810 <TOM>

A/Cross-references: UNIPROT:P12545; GB:J04697; NID:g342272; PIND:AAA36901.1; PID:g342273

C/Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology

C/Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringle; serine proteinase

F/1-9/Domain: signal sequence #status predicted <SIG>

F/103-181/Domain: kringle homology <KR1>

F/185-262/Domain: kringle homology <KR2>

F/275-352/Domain: kringle homology <KR3>

F/377-454/Domain: kringle homology <KR4>

F/481-560/Domain: kringle homology <KR5>

F/581-803/Domain: trypsin homology <TRY>

F/49-73,53-61,103-181,124-164,152-176,185-262,188-316,206-245,234-257,275-352,296-335,324

bonds: #status predicted

F/622,665,760/Active site: His, Asp, Ser #status predicted

Query Match 12.7%; Score 481; DB 2; Length 810;

Best Local Similarity 29.1%, Pred. No. 3.6e-24;  
Matches 159; Conservative 54; Mismatches 169; Indels 164; Gaps 26;

QY 235 EKYC-----GERSOVVTSNSKITYR-----256  
 Db 321 ENYCRNPDGEKAPMCYTNNSQ---VRMEYCKIPSCSSPVSTELDPAPAPLTPVVOEC 377  
 QY 257 FHSQ-QSYTDTGFLAEVLSYDSSDPCEQFTCRTRGCIKRELKRGDGNADCTHSDLNCS 315  
 Db 378 YHGDGQSYRGT-----SSTTTTGK-----KQSSMSMTPH-----407  
 QY 316 CDAGHQTCKRKKFKCKPLFWVCDSDVNDGDNDSDEGSCG---PQTRCSNGKCLSTSQ 371  
 Db 408 ---WHEKTPENFPAAGL-----TNNYCRNPDADGPMCFPTTDPSEVMEYCNLK-----K 453  
 QY 372 CNG-----KDDC---GDGSDASCPKVVNVTC-----KHT 399  
 Db 454 CSGTSGSVAAAPPPVLAQLPDAETPSEBDCMFGNGKGYRG-KKATTVTGTPOCEMAAQPBS 512  
 QY 400 YRCIN-----GL-----CLSK-----GNPEC---DGEKDCSDGDEKDC-----DCGLRSF 437  
 Db 513 HRIFTPETNPRAEGEKVYCRNPDGVGPMCYTNNPKLFDYCDVPCQASSPFCGRQV 572  
 QY 438 TRQ---ARVVGCTADGEMFWQVSLHALGCGHICGASLISPMVLVSAACVYIDRGFRY 494  
 Db 573 EPKCKPGRVVGCVAYPHSWPMQSLSTRRLGMHFCGGTLLSPFWLTAHCL-----EKS 627  
 QY 495 SDPTQMTAFILGHDSQSRAPGVORERLKRILSHPFNDFFVDIALLEKPAEYSSM 554  
 Db 628 SRPFYVILICAH-REVALEPHVEIEVSKMFSR-----ARADILLKLKSLPAIITTDK 680  
 QY 555 VRPILCPDASHVPAGKAIWVTGHTQYGGTGLILQKEIRVYNQTCF---NLPLPOI 612  
 Db 681 VIPACLPSPNVVADRTCEFTTGGGFTQ-GTYGAGLLKEARLPYLENVKVCRRYEFNGTV 739  
 QY 613 TPRMCTGFLSGVDSQCGSGGLSSVEADGRIFQAGVYVSWGCGAQRRKPGVYTLPL 672  
 Db 740 KTELCAAGHLAGTDSQCGSGGLVCFEKDKYILQ-GVTSWIGCARPNKPGVYVSVR 798  
 QY 673 FRDWIK 678  
 Db 799 FVTWIE 804

RESULT 25  
 S33777  
 hepsin (EC 3.4.21.-) - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 16-Aug-2004  
 C/Accession: S33777; S32013  
 R/Parley, D.; Raymond, F.; Nick, H.  
 Biochim. Biophys. Acta 1173, 350-352, 1993  
 A>Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.  
 A:Reference number: S33777; MUID:93305733; PMID:8318546  
 A/Accession: S33777  
 A/Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-416 <FAR>  
 A:Cross-references: UNIPROT:O05511; EMBL:X70900; NID:957928; PIDN:CAA50256.1; PID:957928  
 C:Superfamily: trypsin homology  
 C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein  
 F/22-44/Domain: transmembrane #status predicted <TNN>  
 F/162-399/Domain: transmembrane #status predicted <TRY>  
 F/187-203,290-358,321-337,348-380/Disulfide bonds: #status predicted  
 F/202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 12.7%, Score 480.5; DB 1; Length 416;  
 Best Local Similarity 38.6%; Pred. No. 2e-24;  
 Matches 105; Conservative 42; Mismatches 96; Indels 29; Gaps 10;

QY 431 DCGLRSTQRQARVVGTDADGEMFWQVSLHALGCGHICGASLISPMVLVSAACVYIDDR 490  
 Db 151 DCGRRKLVD-RIVGGQDSLSGRPMQVSLKYDTG-HLCCGSLISGMVLTAAHCPE-- 206

QY 491 GFRYSPTQWTAFLGLHDSQSRAPGVORERLKRILISH-----PFNDFTPD---YDIAL 543  
 Db 207 --ENRVLSTRRVPAG---AAARTSPHAYGVQAVIYHGYLP-FRDPITDENSIDIALV 260  
 QY 544 ELEKPAEYSSMVRPILCPDASHVPAGKAIWVTGHTQYGGTGLILQKEIRVYNQTT 603  
 Db 261 HLSSSLPRTYIOPVCLPAAGALVDGKCTVYVGNQTOYGGQAVVLOEARVPIISNEV 320  
 QY 604 CE---NLPLPOITPRMCMVGLSGVDSQCGSDGAPL---SSVADGRIFQAGVSWMGDC 658  
 Db 321 CNSPDPFGNOKIKRMKCAGYPEGGIDACQDSGCHVCEBRISGTRMRLCGIVSNQTCG 380  
 QY 659 AQRNKPQVYTRLPFRDWI-----KENTGV 683  
 Db 381 ALARKPGVYTKVIDPFRWIFQAIKTHSEATGM 412

RESULT 26  
 KFHUI  
 coagulation factor Xia (EC 3.4.21.27) precursor [validated] - human  
 N/Alternate names: antihemophilic factor C; plasma thromboplastin antecedent  
 C:Species: Homo sapiens (man)  
 C/Date: 13-Aug-1986 #sequence\_revision 26-May-1994 #text\_change 09-Jul-2004  
 C/Accession: A27431; A00920; A37940  
 R/Aakal, R.; Davie, E.W.; Chung, D.W.  
 Biochemistry 26, 7221-7228, 1987  
 A>Title: Organization of the gene for human factor XI.  
 A:Reference number: A27431; MUID:88107663; PMID:2827746  
 A/Accession: A27431  
 A:Molecule type: DNA  
 A:Residues: 1-625 <ASA>  
 A:Cross-references: UNIPROT:P03951; GB:M18295  
 A/Note: the sequence shown follows the authors' translation  
 R/Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.  
 Biochemistry 25, 2417-2424, 1986  
 A>Title: Amino acid sequence of human factor XI, a blood coagulation factor with four car  
 A:Reference number: A00920; MUID:86243360; PMID:3636155  
 A/Accession: A00920  
 A:Molecule type: mRNA  
 A:Residues: 1-625 <FLU>  
 A:Cross-references: GB:M13142; NID:g182832; PIDN:AAA52487.1; PID:g182833  
 R/McMullen, B.A.; Fujikawa, K.; Davie, E.W.  
 Biochemistry 30, 2056-2060, 1991  
 A/Title: Location of the disulfide bonds in human coagulation factor XI: the presence of  
 A:Reference number: A37940; MUID:91152017; PMID:1998667  
 A/Accession: A37940  
 A:Molecule type: protein  
 A:Residues: 28-33/35-49, 'X', 51-55, 'X', 57-63/70-75, 'X', 77-79/107-109, 'X', 111-112/132-139, 'X', 1280-282, 'X', 284/285-297/313-316, 'X', 318-319/320-326, 'X', 328-330, 'X', 347-349/373, 'X', 375;  
 C/Comment: The proenzyme consists of two identical chains linked by one or more disulfide  
 he active site, and a heavy chain, which associates with high molecular weight (HMW) kin  
 C:Genetics:  
 A:Gene: GDB:F11  
 A:Cross-references: GDB:119891; OMIM:264900  
 A/Map position: 4q35-4q35  
 A/Intons: 19/1; 73/2; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1; 526  
 C/Function:  
 A>Description: catalyzes the proteolytic activation of coagulation factor IX  
 A/Pathway: blood coagulation intrinsic pathway  
 C:Superfamily: coagulation factor XI; trypsin homology  
 C:Keywords: blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hydro  
 F/1-18/Domain: signal sequence #status predicted <SIG>  
 F/19-387/Product: coagulation factor Xia heavy chain #status experimental <HCH>  
 F/19-108/Domain: apple repeat <AP1>  
 F/109-198/Domain: apple repeat <AP2>  
 F/199-288/Domain: apple repeat <AP3>  
 F/290-379/Domain: apple repeat <AP4>  
 F/388-625/Product: coagulation factor Xia light chain #status experimental <LCH>  
 F/388-618/Domain: trypsin homology <TRY>  
 F/20-103,514-581,571-599/Disulfide bonds: #status predicted  
 F/22/Disulfide bonds: interchain #status experimental  
 F/46-76,50-56,110-193,136-165,140-146,200-283,226-255,230-236,291-374,317-346,321-327,380



F;72-90,142-156,152-165,167-180,184-211,241-259,300-348,328-361,366-412,396-430,434-552,  
F;158/Modified site: erythro-beta-hydroxyasparagine (asn) #status predicted  
F;444-445/Cleavage site: Arg-Ile (autolytic) #status predicted  
F;483,532,633/Active site: His, Asp, Ser #status predicted

Query Match 12.5%; Score 474; DB 1; Length 686;  
Best Local Similarity 25.3%; Pred. No. 8,9e-24;  
Matches 184; Conservative 89; Mismatches 275; Indels 180; Gaps 30;

```

55 RFTTPGPPDPYPHARCOMALRGDADSVLSLTFPSRDLASCDBRSGDLVTVVNTLS PME 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db RLASPGFP-GEYANDQRRMTLTAPPGYRLRLVPTTHDLBSLCEYDFVTLSSGAA- 84
QY 115 PHALVOLCG-----TTPPSYNTLTFHSSQNVLLITLITNTERRP--GFEATFPQLPRMS 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 --VLATLGGQESTTERAPGKDTFYSLGSSLDITFRSDYSEKPTGFPA-FVAEDIDE 141
QY 168 C-----GGLRKAQGTFRNPPY 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142 QVAPGAPTCDDHCHNLGGFYCSGAGVYLHNKRTCSALCSGQVFTORSGELSP EY 201
QY 185 PGHYPPIIDCTWNIEVRNNGHVKVPKPFYLLBPRACGCPKDYVINGEK-----YCG 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 PRPFPKLSCTYSISLEBGSVILDFVESPVVEHPET-LCPYDFLKIQTDRBHGPF CG 260
QY 240 ERSQFVVTNSNKITVRFHSDQSYTDGFLAEYLSYDSDPGP-----GQFTCRGR CIR 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 KTLPHRIETKSNVTITFTVDESQDHTGMKIHVSTHACPYMAPRNGHVSPIVQAKYIL 320
QY 295 KEALRCDMDACTHSDDELNCSQAGHOFCTCKNCKEFLFWVCDSVNDCGNSDBO---GC 351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 K-----DSFSIFCETGYELQGHLPKSFVAVCK-----DGSMDRPM PAC 361
QY 352 SCPAOTFRCSNGKCLSKSQCCNGKDCGDSDEASCPRVAVVTCCKHT-----YRCLN 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 SI V-----DCGP-PDILPSGRVEYITPGYTTKAVI QYCEB 398
QY 405 GLCLSKGNPECDGKEDC-SDG-----SDEKD---CD--CGLRSTFRQARVVGTDADGE 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 TFFYTMKN---DKKYCEADGFMFTSKGKSLPVCBPCGHSATTTGRIIGGQKARPD 455
QY 454 WPNQVSLHALGCGHICGASLISPMWLVSAAHCYIDDDGFYSPTQMTAFLGLHDSQRS 513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 FPMQVLI--LG-GTTAAGALLYDMWVLTAAHAYEOK-----HD--AS 493
QY 514 APGOERELKRIISH-----PFNDFTPYDIALLELEKPAEYSSWVRPICL 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 494 ALDIRMGTLRKLSHYTOAMSSEAVFHIEGTHDAGFNDIALITLKNKVINSNITPICL 553
QY 561 P--DASHVPPAGKAIWVTGMHTQYGTGALLQKGEIRVINOTTGENTLLPQ-----QI 612
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 554 PRKEABSFMRFTDIDGTASGMLTDRGFL-ARNLMYVIVIPYDHQKCTAAVEKPPYPRGSV 612
QY 613 TPRMCCGPGISGVDSGCGSGGSLSSVEAD-GRIPQAGVSWMD-CCAQNRKGVYTRL 670
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 613 TANNLCGLBSGKSGDSGRSGGALVFLDSETERWFVGIGIVSGSNMGSGAGQGYVTKV 672
QY 671 PLFRDWIK 678
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 673 INVIPWIE 680

```

# RESULT 29

S40007  
trypsin (BC 3.4.21.4) precursor - African malaria mosquito  
C:Species: Anopheles gambiae (African malaria mosquito)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
C/Accession: S40007  
R:Mueller, H.  
submitted to the EMBL Data Library, June 1993  
A:Reference number: S40003  
A/Accession: S40007  
A>Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-275 <MBE>  
A:Cross-references: UNIPROT:P35037; EMBL:Z22930; NID:G410323; PIDN:CAA80517.1; PID:G4103;  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; protein digestion; serine proteinase  
F;48-269/Domain: trypsin homology <TRY>  
F;89,134,230/Active site: His, Asp, Ser #status predicted

Query Match 12.5%; Score 472; DB 2; Length 275;  
Best Local Similarity 39.9%; Pred. No. 4,8e-24;  
Matches 97; Conservative 39; Mismatches 89; Indels 18; Gaps 6;

```

QY 442 RVVGTDADEGEPMQVSHALGQGHICGASLISPMWLVSAAHCYIDDDRGFRYSDPTQWT 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 RIVGGEFIDVSETPYVVSQYF-NSHRCGSVLNSKMIILPAACVNA-----LQPSLA 100
QY 502 AFLGLHDSQRSAPGOERELKRIISHPPNDFTPYDIALLELEKPAEYSSWVRPICL 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 VRLG---SSRHASGTVVAVARVLEHFNVDSTIDYDSLMELETBLTFSDVQVSLP 156
QY 562 DASHVPPAGKAIWVTGMHTQYGTGALLQKGEIRVINOTTGENTLLPQ--ITPRMVCV 619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 157 EQDEAVEDGIMTVSGMGNTQSAEASNALIRANIFTVNQKECTIAYSSGGITDMLCA 216
QY 620 GFLSGVDSGQDSGCPPLSSVEADGRIPQAGVSWEDGCAQRNKPQYTRLPPLFRDWIK 679
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 217 GYRGGKDALQCGDSGGL---VVDGRL--VGVSWSGFGCAMPQPGVYAVAVRVMVRE 271
QY 680 NTG 682
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 NSG 274

```

## RESULT 30

A61545  
Plasmin (BC 3.4.21.7) precursor - horse (fragments)  
N:Alternate names: plasminogen  
N:Contains: miniplasminogen  
C:Species: Equus caballus (domestic horse)  
C>Date: 28-Oct-1994 #sequence\_revision 01-Nov-1996 #text\_change 09-Jul-2004  
C/Accession: A61545; PMID:8905015; PMID:3168975  
R:Schaller, J.; Rickli, E.E.  
Enzyme 40, 63-69, 1988  
A:Title: Structural aspects of the plasminogen of various species.  
A:Reference number: A61545; PMID:8905015; PMID:3168975  
A/Accession: A61545

A:Molecule type: protein

A:Residues: 1-33,34-117 <SCH>

A:Cross-references: UNIPROT:P80010

R:Schaller, J.; Straub, C.; Kaempfer, U.; Rickli, E.E.

Protein Seq. Data Anal. 4, 69-74, 1991

A:Title: Complete amino acid sequence of equine miniplasminogen.

A:Reference number: S17527; PMID:92052077; PMID:1946332

A/Accession: S17527

A:Molecule type: protein

A:Residues: 118-455 <SC2>

C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology;

C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; plasma; serine proteinase; z

F;1-33,34-117,118-455/Product: plasminogen (fragments) #status experimental <PRO>

F;1-33/Domain: activation peptide (fragment) #status experimental <APR>

F;34-117,118-225,226-455/Product: plasmin (fragments) #status experimental <MAT>

F;37-114/Domain: kringle homology <KR4>

F;118-455/Product: miniplasminogen #status experimental <MIN>

F;126-205/Domain: kringle homology <KR5>

F;226-455/Domain: plasmin chain B #status experimental <BCH>

F;267-448/Domain: trypsin homology <TRY>

F;267,310,405/Active site: His, Asp, Ser #status predicted

Query Match 12.5%; Score 472; DB 2; Length 455;  
Best Local Similarity 31.9%; Pred. No. 8e-24;  
Matches 138; Conservative 49; Mismatches 155; Indels 90; Gaps 19;

QY 298 RCDGMADCTDHSDELNCSQAGHOFCTCKNCKEFLFWVCDSVNDCGNSDBDQGCSC----- 353

```

Db      57  KQOSWSSWTPH-----MHQKTPREKYPNADL-----TMNYCRNPDGDKGPMCYTTD 101
Qy      354 PAQTFR-CSNGKCLSKSQCCNGKDDC---GDGSDASCPRKYNVTCT-----KHTY 400
Db      102 PSVMEFCNLKRC-SETVQEPSEBPCMGIGKYGKKATTVTGTRCQAMAAQEBHRHSI 160
Qy      401 -----RC-----LNGLCLSKNPB-----CDQKE-----DCSDGSDDE-K 428
Db      161 FTPEANPEANLEKRYCERNPDGDVNGPWCYTMNPKLPDYCDVPOCESSPFDCGKRYEPK 220
Qy      429 DCDGCLRSFTRQARVVGSTDADEGEWPQVSLHALGCGHICGASLISPMLVSAHCYID 488
Db      221 KC-----SGRIVGGCVAILAHSMFWQISLRTRFGRHFCGTLISPFWTLTAHCL-- 269
Qy      489 DRGRYSDPTQWTFALGLHDQSORSAPGVQERRLKRIISHPFENDFTDYDIALLELEK 548
Db      270 ---ERSSRPSTYKVLGTNHE-LRLAAGAOQIDVSKLFLER-----SRADIALKLSSP 319
Qy      549 AEYSMWPRICLPDASHVFPAGKAIWYWGHTOYCGTGALILQKGEIRVINQTTCE--N 606
Db      320 AITQWYIPACLPADYVAVANMAECFTVGMGETO-DSSNAGVLKEAQLPVIENKVCNRYE 378
Qy      607 LLPQOITPRMWCVFLSGVDSCQDSCGPLSVADGRIFOAGVSWGDCGQBNKPGV 666
Db      379 YLNGRVXSTELCAGHLVGVVSCQDSCGPLVCFEKDKYILQ-GVTSWGLGCAHPNKGV 437
Qy      667 YTRLPLFRDWIK 678
Db      438 YVRVSSFINWIE 449

```

Search completed: November 29, 2004, 08:33:22  
 Job time : 30.3127 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:32:39 / Search time 88.8166 Seconds

(without alignments)  
2727.603 Million cell updates/sec

Title: US\_09-936-333-5

Perfect score: 3789

Sequence: 1 MABERVVLMPPRRARSLKSFV.....PGVYTRLPFRWDIKENTGV 683

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1575965 seqs, 354694765 residues

Total number of hits satisfying chosen parameters: 1575965

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/prodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/prodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/prodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/prodata/2/pubppaa/PCUTUS\_PUBCOMB.pep:\*  
6: /cgn2\_6/prodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
7: /cgn2\_6/prodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
8: /cgn2\_6/prodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
9: /cgn2\_6/prodata/2/pubppaa/US09B\_PUBCOMB.pep:\*  
10: /cgn2\_6/prodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
11: /cgn2\_6/prodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/prodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
13: /cgn2\_6/prodata/2/pubppaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/prodata/2/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/prodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/prodata/2/pubppaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/prodata/2/pubppaa/US10E\_NEW\_PUB.pep:\*  
18: /cgn2\_6/prodata/2/pubppaa/US10F\_NEW\_PUB.pep:\*  
19: /cgn2\_6/prodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/prodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3760	99.2	762	US-10-729-807-1	Sequence 1, Appl1
2	3760	99.2	855	US-10-295-027-1185	Sequence 1185, Ap
3	3760	99.2	855	US-10-072-012-353	Sequence 353, App
4	3760	99.2	855	US-10-072-012-412	Sequence 412, App
5	3760	99.2	855	US-10-072-012-419	Sequence 419, App
6	3756	99.1	855	US-10-072-012-354	Sequence 354, App
7	3756	99.1	855	US-10-072-012-420	Sequence 420, App
8	3756	99.1	855	US-10-037-417-132	Sequence 132, App
9	3754	99.1	757	US-10-072-012-44	Sequence 44, Appl1
10	3754	99.1	855	US-09-776-191-2	Sequence 2, Appl1
11	3754	99.1	855	US-10-099-700A-2	Sequence 2, Appl1
12	3754	99.1	855	US-10-190-030B-2	Sequence 2, Appl1
13	3754	99.1	855	US-10-302-840A-2	Sequence 2, Appl1

14	3754	99.1	855	US-10-267-219-2	Sequence 2, Appl1
15	3754	99.1	855	US-10-112-221A-2	Sequence 2, Appl1
16	3754	99.1	855	US-10-104-271-2	Sequence 2, Appl1
17	3754	99.1	855	US-10-147-211A-2	Sequence 2, Appl1
18	3754	99.1	855	US-10-156-214A-2	Sequence 2, Appl1
19	3754	99.1	855	US-10-072-012-352	Sequence 352, App
20	3754	99.1	855	US-10-072-012-411	Sequence 411, App
21	3754	99.1	855	US-10-072-012-418	Sequence 418, App
22	3754	99.1	855	US-10-600-187-2	Sequence 2, Appl1
23	3745	98.8	851	US-10-276-774-1798	Sequence 1143, Ap
24	3745	98.8	851	US-10-296-115-1143	Sequence 1143, Ap
25	3259.5	86.0	782	US-10-097-340-312	Sequence 312, App
26	3209	84.7	855	US-09-900-751-2	Sequence 2, Appl1
27	3209	84.7	855	US-10-072-012-355	Sequence 355, App
28	3197	84.4	855	US-10-072-012-413	Sequence 413, App
29	3197	84.4	855	US-10-072-012-356	Sequence 356, App
30	3197	84.4	855	US-10-072-012-414	Sequence 414, App
31	3197	84.4	855	US-10-072-012-417	Sequence 417, App
32	3123	82.4	902	US-10-333-743-3	Sequence 3, Appl1
33	3123	82.4	902	US-10-600-187-10	Sequence 10, Appl1
34	3123	82.4	902	US-10-297-987B-11	Sequence 11, Appl1
35	2797	73.8	620	US-09-925-301-1193	Sequence 1193, Ap
36	2254.5	59.5	845	US-10-072-012-415	Sequence 415, App
37	1319	34.8	241	US-09-776-191-50	Sequence 50, Appl1
38	1319	34.8	241	US-10-099-700A-4	Sequence 4, Appl1
39	1319	34.8	241	US-10-092-004A-2	Sequence 2, Appl1
40	1319	34.8	241	US-10-190-030B-4	Sequence 4, Appl1
41	1319	34.8	241	US-10-302-840A-4	Sequence 4, Appl1
42	1319	34.8	241	US-10-267-219-4	Sequence 4, Appl1
43	1319	34.8	241	US-10-112-221A-4	Sequence 4, Appl1
44	1319	34.8	241	US-10-104-271-4	Sequence 4, Appl1
45	1319	34.8	241	US-10-147-211A-4	Sequence 4, Appl1
46	1029	27.2	850	US-09-888-615-108	Sequence 108, App
47	1019.5	26.9	833	US-10-333-743-7	Sequence 7, Appl1
48	1014	26.8	799	US-10-072-012-410	Sequence 410, App
49	1014	26.8	799	US-10-072-012-416	Sequence 416, App
50	1005	26.5	802	US-09-888-615-113	Sequence 113, App
51	1005	26.5	802	US-09-978-295A-169	Sequence 169, App
52	1005	26.5	802	US-09-978-697-169	Sequence 169, App
53	1005	26.5	802	US-09-978-192A-169	Sequence 169, App
54	1005	26.5	802	US-09-999-832A-169	Sequence 169, App
55	1005	26.5	802	US-09-978-189-169	Sequence 169, App
56	1005	26.5	802	US-09-978-608A-169	Sequence 169, App
57	1005	26.5	802	US-09-978-585A-169	Sequence 169, App
58	1005	26.5	802	US-09-978-191A-169	Sequence 169, App
59	1005	26.5	802	US-09-978-403A-169	Sequence 169, App
60	1005	26.5	802	US-09-978-564A-169	Sequence 169, App
61	1005	26.5	802	US-09-999-833A-169	Sequence 169, App
62	1005	26.5	802	US-09-981-915A-169	Sequence 169, App
63	1005	26.5	802	US-09-978-824-169	Sequence 169, App
64	1005	26.5	802	US-09-978-585A-169	Sequence 169, App
65	1005	26.5	802	US-09-999-834A-169	Sequence 169, App
66	1005	26.5	802	US-09-978-423A-169	Sequence 169, App
67	1005	26.5	802	US-09-978-193A-169	Sequence 169, App
68	1005	26.5	802	US-09-999-830A-169	Sequence 169, App
69	1005	26.5	802	US-09-978-757A-169	Sequence 169, App
70	1005	26.5	802	US-09-978-187B-169	Sequence 169, App
71	1005	26.5	802	US-09-978-643A-169	Sequence 169, App
72	1005	26.5	802	US-09-978-375A-169	Sequence 169, App
73	1005	26.5	802	US-09-978-298A-169	Sequence 169, App
74	1005	26.5	802	US-09-978-188A-169	Sequence 169, App
75	1005	26.5	802	US-09-978-681A-169	Sequence 169, App
76	1005	26.5	802	US-09-999-828A-169	Sequence 169, App
77	1005	26.5	802	US-09-999-829A-169	Sequence 169, App
78	1005	26.5	802	US-09-978-544A-169	Sequence 169, App
79	1005	26.5	802	US-09-978-665A-169	Sequence 169, App
80	1005	26.5	802	US-09-978-802A-169	Sequence 169, App
81	1005	26.5	802	US-09-999-831A-169	Sequence 169, App
82	1005	26.5	802	US-10-017-081A-169	Sequence 169, App
83	1005	26.5	802	US-10-167-749-169	Sequence 169, App
84	1005	26.5	802	US-10-013-921A-169	Sequence 169, App
85	1005	26.5	802	US-10-013-929A-169	Sequence 169, App
86	1005	26.5	802	US-10-013-929A-169	Sequence 169, App



```
87 1005 26.5 802 14 US-10-016-177A-169 Sequence 169, App
88 1005 26.5 802 14 US-10-166-709A-169 Sequence 169, App
89 1005 26.5 802 14 US-10-143-031A-169 Sequence 169, App
90 1005 26.5 802 14 US-10-143-030A-169 Sequence 169, App
91 1005 26.5 802 14 US-10-002-967A-169 Sequence 169, App
92 1005 26.5 802 14 US-10-017-083A-169 Sequence 169, App
93 1005 26.5 802 14 US-10-145-128A-169 Sequence 169, App
94 1005 26.5 802 14 US-10-017-191A-169 Sequence 169, App
95 1005 26.5 802 14 US-10-143-028A-169 Sequence 169, App
96 1005 26.5 802 14 US-10-143-029A-169 Sequence 169, App
97 1005 26.5 802 14 US-10-145-089A-169 Sequence 169, App
98 1005 26.5 802 14 US-10-165-067A-169 Sequence 169, App
99 1005 26.5 802 14 US-10-145-017A-169 Sequence 169, App
100 1005 26.5 802 14 US-10-164-728A-169 Sequence 169, App
```

## ALIGNMENTS

```
RESULT 1
; Sequence 1, Application US/10729807
; Publication No. US20040132158A1
; GENERAL INFORMATION:
; APPLICANT: BANDMAN, Olga; HILLMAN, Jennifer L.
; APPLICANT: TANG, Y. Tom; LAL, Preeti G.
; APPLICANT: YUE, Henry; AZIMZAI, Valda
; APPLICANT: BAUGHN, Mariah R.; LU, Dyung Aina M.
; TITLE OF INVENTION: HUMAN PEPTIDASES
; FILE REFERENCE: PF-0651-1 DIV
; CURRENT APPLICATION NUMBER: US/10/729, 807
; CURRENT FILING DATE: 2003-12-05
; PRIOR APPLICATION NUMBER: US 09/889, 238
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: PCT/US00/00641
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: US 60/172, 247
; PRIOR FILING DATE: 1999-01-11
; PRIOR APPLICATION NUMBER: US 60/132, 253
; PRIOR FILING DATE: 1999-05-03
; PRIOR APPLICATION NUMBER: US 60/136, 653
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 155179CD1
US-10-729-807-1
```

```
Query Match 99.2%; Score 3760; DB 16; Length 762;
Best Local Similarity 99.4%; Pred. No. 2e-275;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 MAERVVMLPPRRSLKSFVVTGVVAFPTDSTKYORTODNSCGFGLHARGVELMRFTTPG 60
DB 80 MAERVVMLPPRRSLKSFVVTGVVAFPTDSTKYORTODNSCGFGLHARGVELMRFTTPG 139
QY 61 PPDSPYAHARCOMALRGDADSVLSLTFRSFDLASCDERSGDLVTVNTLSPEMHALVQ 120
DB 140 PPDSPYAHARCOMALRGDADSVLSLTFRSFDLASCDERSGDLVTVNTLSPEMHALVQ 199
QY 121 LCGTTPSYNLTTHSSQNVLLITLITNTERRHGFEATFQQLPRMSSCGRLKRAQGTEN 180
DB 200 LCGTTPSYNLTTHSSQNVLLITLITNTERRHGFEATFQQLPRMSSCGRLKRAQGTEN 259
QY 181 SPYYPGYPNICTNWIEVNNOHVVRKFPFLLEBRACGTCPCDVYVEINGEKYCGE 240
DB 260 SPYYPGYPNICTNWIEVNNOHVVRKFPFLLEBRACGTCPCDVYVEINGEKYCGE 319
```

```
QY 241 RSQFVVTNSNKTITVFHSDOSYTDGFLAFLYLSYDSDPCPGQFTCRTRCIRKELRCD 300
DB 320 RSQFVVTNSNKTITVFHSDOSYTDGFLAFLYLSYDSDPCPGQFTCRTRCIRKELRCD 379
QY 301 GMAICTDHSDELNCSDAGHQFTCKNKFCKPLFWVCDVNDGDNDEQCCSPAQTFR 360
DB 380 GMAICTDHSDELNCSDAGHQFTCKNKFCKPLFWVCDVNDGDNDEQCCSPAQTFR 439
QY 361 SNKCLSKSQCGKXGKDCGSDGDEASC PKNVVYCTKHTYRCLNGCLSKGNECCKED 420
DB 440 SNKCLSKSQCGKXGKDCGSDGDEASC PKNVVYCTKHTYRCLNGCLSKGNECCKED 499
QY 421 CSDGSEKDCDCLRSFTTQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480
DB 500 CSDGSEKDCDCLRSFTTQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMVLV 559
QY 481 SAACHYIDRGRFYSDPTQWTAFLGLHDOSQRAAPGVQERRLRRIISHPPNFTFDYDI 540
DB 560 SAACHYIDRGRFYSDPTQWTAFLGLHDOSQRAAPGVQERRLRRIISHPPNFTFDYDI 619
QY 541 ALLELEKPAEYSMWBPICLPDASHVFPAGKAIWVTGMGHTGYGTAFLQKGEIRVIN 600
DB 620 ALLELEKPAEYSMWBPICLPDASHVFPAGKAIWVTGMGHTGYGTAFLQKGEIRVIN 679
QY 601 QTTCEMLLPQOITPRMNCVGFSLGVDSCQGSBGFLSSVEADGRIFOAGVSWGDCAQ 660
DB 680 QTTCEMLLPQOITPRMNCVGFSLGVDSCQGSBGFLSSVEADGRIFOAGVSWGDCAQ 729
QY 661 RNRPGVYTRLPFLFRDWIKENTGV 683
DB 740 RNRPGVYTRLPFLFRDWIKENTGV 762
```

```
RESULT 2
US-10-295-027-1185
; Sequence 1185, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natsheha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295, 027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663, 733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350, 666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335, 394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332, 464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334, 393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/347, 349
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/355, 250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356, 714
; PRIOR FILING DATE: 2002-02-13
```

```

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 1185
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1185

Query Match      99.2%; Score 3760; DB 14; Length 855;
Best Local Similarity 99.4%; Pred. No. 2.2e-275;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRASLSFVVTSVVAPPTDSTKYQRTODNSCGFLHARGVELMRFTTPG 60
DB 173 MAERVVMLPPRASLSFVVTSVVAPPTDSTKYQRTODNSCGFLHARGVELMRFTTPG 232
QY 61 FPDSPYPAHARCOMALRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVQ 120
DB 233 FPDSPYPAHARCOMALRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVQ 292
QY 121 LCGTTPPSYNLTFRSSQNVLLITLITNTERRHGFATFPQLPRMSSCGRLRAOQTFN 180
DB 293 LCGTTPPSYNLTFRSSQNVLLITLITNTERRHGFATFPQLPRMSSCGRLRAOQTFN 352
QY 181 SPYRGHYPPNIDCTNMIIEVNNQHVVRKFFYLLEPRACGTCPCDYIENGKTCGE 240
DB 353 SPYRGHYPPNIDCTNMIIEVNNQHVVRKFFYLLEPRACGTCPCDYIENGKTCGE 412
QY 241 RSQFVVTNSNKTIVRFHSDQSYTDGTGFLAEVLSYDSSDPCPGQFTCRGTCIRKEIRCD 300
DB 413 RSQFVVTNSNKTIVRFHSDQSYTDGTGFLAEVLSYDSSDPCPGQFTCRGTCIRKEIRCD 472
QY 301 GMADCTHSDBLNCSCDAGHQTCKNKECKPLFWVCDSVNDGNSDEOGSCSPAQTFRC 360
DB 473 GMADCTHSDBLNCSCDAGHQTCKNKECKPLFWVCDSVNDGNSDEOGSCSPAQTFRC 532
QY 361 SNGCCLSKSQOCCNGKDDCGSDSDASCPKNNVVTCTKTTYCLNGLCLSKNPNCDGED 420
DB 533 SNGCCLSKSQOCCNGKDDCGSDSDASCPKNNVVTCTKTTYCLNGLCLSKNPNCDGED 592
QY 421 CSDSDDEKDCDCCGRLSFTROARVVGTDADGEMPMQVSLALQGHICGSLISPMNLV 480
DB 593 CSDSDDEKDCDCCGRLSFTROARVVGTDADGEMPMQVSLALQGHICGSLISPMNLV 652
QY 481 SAACHYIDDRGRYSDDPTQWTAFLGLHDQSORSAAPGVBERLKRILISHPFNDFTFDYDI 540
DB 653 SAACHYIDDRGRYSDDPTQWTAFLGLHDQSORSAAPGVBERLKRILISHPFNDFTFDYDI 712
QY 541 ALLELEKPAEYSSWVRPCLPDASHVPPAKAIWVTGHTQYGGTGALLIQKEIRIYN 600
DB 713 ALLELEKPAEYSSWVRPCLPDASHVPPAKAIWVTGHTQYGGTGALLIQKEIRIYN 772
QY 601 QTTBENLLPQOITPRMNCVGLSGSVDSCGSDSGGFLSSVADRIIFQAGVNSGDCAQ 660
DB 773 QTTBENLLPQOITPRMNCVGLSGSVDSCGSDSGGFLSSVADRIIFQAGVNSGDCAQ 832
QY 661 RNKPGVYTRLPPLFRDWIKENTGV 683
DB 833 RNKPGVYTRLPPLFRDWIKENTGV 855

RESULT 3
US-10-072-012-353
; Sequence 353, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tcheu, Veilzar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zeehuysen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shimkete, Richard
; APPLICANT: Li, Li

; APPLICANT: Gangoli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsebrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 353
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-353

Query Match      99.2%; Score 3760; DB 15; Length 855;
Best Local Similarity 99.4%; Pred. No. 2.2e-275;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRASLSFVVTSVVAPPTDSTKYQRTODNSCGFLHARGVELMRFTTPG 60
DB 173 MAERVVMLPPRASLSFVVTSVVAPPTDSTKYQRTODNSCGFLHARGVELMRFTTPG 232
QY 61 FPDSPYPAHARCOMALRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVQ 120
DB 233 FPDSPYPAHARCOMALRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVQ 292
QY 121 LCGTTPPSYNLTFRSSQNVLLITLITNTERRHGFATFPQLPRMSSCGRLRAOQTFN 180
DB 293 LCGTTPPSYNLTFRSSQNVLLITLITNTERRHGFATFPQLPRMSSCGRLRAOQTFN 352
QY 181 SPYRGHYPPNIDCTNMIIEVNNQHVVRKFFYLLEPRACGTCPCDYIENGKTCGE 240
DB 353 SPYRGHYPPNIDCTNMIIEVNNQHVVRKFFYLLEPRACGTCPCDYIENGKTCGE 412
QY 241 RSQFVVTNSNKTIVRFHSDQSYTDGTGFLAEVLSYDSSDPCPGQFTCRGTCIRKEIRCD 300
DB 413 RSQFVVTNSNKTIVRFHSDQSYTDGTGFLAEVLSYDSSDPCPGQFTCRGTCIRKEIRCD 472
```

```
QY 301 GMADCTDHSDELNCS CDAGHQFTCKNFKCKPLFWVCDSVND CGDNDSDGSCSPAQTFRC 360
D 473 GMADCTDHSDELNCS CDAGHQFTCKNFKCKPLFWVCDSVND CGDNDSDGSCSPAQTFRC 532
QY 361 SNGKCLSKSQOQCNKGKDDCGDSDASCPKXNVVTCYKHYRCLNGLCLSKGNPCCDCKED 420
D 533 SNGKCLSKSQOQCNKGKDDCGDSDASCPKXNVVTCYKHYRCLNGLCLSKGNPCCDCKED 592
QY 421 CSDGSDKDCDCGLRSFTROARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNMLV 480
D 593 CSDGSDKDCDCGLRSFTROARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNMLV 652
QY 481 SAHCYIDDRGFRYSPTQWTAFLGLHDQSRSAPGVQERLRKRIISHPEFNDFTDYDI 540
D 653 SAHCYIDDRGFRYSPTQWTAFLGLHDQSRSAPGVQERLRKRIISHPEFNDFTDYDI 712
QY 541 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGHTQYCGTGALLIQKEIRVIN 600
D 713 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGHTQYCGTGALLIQKEIRVIN 772
QY 601 QTTCEMLPQOITPRMNCVGLSGVDS CGDSCGFLSSVEADGRIFQAGVSWGDCAQ 660
D 773 QTTCEMLPQOITPRMNCVGLSGVDS CGDSCGFLSSVEADGRIFQAGVSWGDCAQ 832
QY 661 RNKPGVYTRLPFLFRDWIKENTGV 683
D 833 RNKPGVYTRLPFLFRDWIKENTGV 855
```

RESULT 4  
US-10-072-012-412

```
; Sequence 412, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernyev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimiro Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Putrak, Katarzyna
; APPLICANT: Groose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
```

```
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 412
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-412
```

Query Match 99.4%; Score 3760; DB 15; Length 855;  
Best Local Similarity 99.4%; Pred. No. 2,2e-275;  
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 1 MAERVVMLEPPRARSLSKSPVYVSVAFPDSDKTQRTQDNDSCSFGJHARGVELMRFTTPG 60
D 173 MAERVVMLEPPRARSLSKSPVYVSVAFPDSDKTQRTQDNDSCSFGJHARGVELMRFTTPG 232
QY 61 FPDSPYPAHARCOMALRGADSVLSLTPRSFDLASCDERGSDLVTYNTLSPEMHALVQ 120
D 233 FPDSPYPAHARCOMALRGADSVLSLTPRSFDLASCDERGSDLVTYNTLSPEMHALVQ 292
QY 121 LCCTYPPSYNLTFHSSQNLTLTLINTEPRHNGEATFPQLPRMSSCGRLPKAGCTN 180
D 293 LCCTYPPSYNLTFHSSQNLTLTLINTEPRHNGEATFPQLPRMSSCGRLPKAGCTN 352
QY 181 SPYPGHYPPNIDCTNNIEVPPNNOHYKRFKFFYLLEPPRACCTCKDYIEINGEYKCE 240
D 353 SPYPGHYPPNIDCTNNIEVPPNNOHYKRFKFFYLLEPPRACCTCKDYIEINGEYKCE 412
QY 241 RSQFVVTSSNKKITVAFHSDQSYTDGFLAELYSYDSDPCPGQFTCRGCRKELRCD 300
D 413 RSQFVVTSSNKKITVAFHSDQSYTDGFLAELYSYDSDPCPGQFTCRGCRKELRCD 472
QY 301 GMADCTDHSDELNCS CDAGHQFTCKNFKCKPLFWVCDSVND CGDNDSDGSCSPAQTFRC 360
D 473 GMADCTDHSDELNCS CDAGHQFTCKNFKCKPLFWVCDSVND CGDNDSDGSCSPAQTFRC 532
QY 361 SNGKCLSKSQOQCNKGKDDCGDSDASCPKXNVVTCYKHYRCLNGLCLSKGNPCCDCKED 420
D 533 SNGKCLSKSQOQCNKGKDDCGDSDASCPKXNVVTCYKHYRCLNGLCLSKGNPCCDCKED 592
QY 421 CSDGSDKDCDCGLRSFTROARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNMLV 480
D 593 CSDGSDKDCDCGLRSFTROARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNMLV 652
QY 481 SAHCYIDDRGFRYSPTQWTAFLGLHDQSRSAPGVQERLRKRIISHPEFNDFTDYDI 540
D 653 SAHCYIDDRGFRYSPTQWTAFLGLHDQSRSAPGVQERLRKRIISHPEFNDFTDYDI 712
QY 541 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGHTQYCGTGALLIQKEIRVIN 600
D 713 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGHTQYCGTGALLIQKEIRVIN 772
QY 601 QTTCEMLPQOITPRMNCVGLSGVDS CGDSCGFLSSVEADGRIFQAGVSWGDCAQ 660
D 773 QTTCEMLPQOITPRMNCVGLSGVDS CGDSCGFLSSVEADGRIFQAGVSWGDCAQ 832
QY 661 RNKPGVYTRLPFLFRDWIKENTGV 683
D 833 RNKPGVYTRLPFLFRDWIKENTGV 855
```

RESULT 5  
US-10-072-012-419

```
Sequence 419, Application US/10072012
Publication No. US2004003493A1
GENERAL INFORMATION:
APPLICANT: Tchernev, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zerhusen, Bryan
APPLICANT: Patuturajan, Meera
APPLICANT: Shimkets, Richard
APPLICANT: Li, Li
APPLICANT: Gangoli, Beha
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie E.
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimyr Y.
APPLICANT: Colman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Putrak, Katarzyna
APPLICANT: Groese, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072.012
PRIOR FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 419
LENGTH: 855
TYPE: PRT
ORGANISM: Homo sapiens
US-10-072-012-419
Query Match 99.2%; Score 3760; DB 15; Length 855;
Best Local Similarity 99.4%; Pred. No. 2,2e-275;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Db 293 LCGTTPSYNLTTFHSSQNVLLITLITWERRHHPGFATFPQLPRMSSCGRLRAQGTFN 352
Qy 181 SPYPGHYPPNIDCTNNAIEVNNQHVVRKFPYLLEPRRACGTCPDYVEINGEKYCGE 240
Db 353 SPYTPGHYPPNIDCTNNAIEVNNQHVVRKFPYLLEPRRACGTCPDYVEINGEKYCGE 412
Qy 241 RSQFVVTNSNKTITVRHSDQSYTDGFLAELVLSYSSDPCPOQFTCRTRCIRKELRCD 300
Db 413 RSQFVVTNSNKTITVRHSDQSYTDGFLAELVLSYSSDPCPOQFTCRTRCIRKELRCD 472
Qy 301 GMADCTDHSDELNCSCDAGHFTCKNFKCPLFWVCDVNDGDNDEQCSCPAQTRFC 360
Db 473 GMADCTDHSDELNCSCDAGHFTCKNFKCPLFWVCDVNDGDNDEQCSCPAQTRFC 532
Qy 361 SNGKCLSKSQQCKGKDCGSGSDEASCPKNNVYTCRHTRCNLGLCLSGNRECKED 420
Db 533 SNGKCLSKSQQCKGKDCGSGSDEASCPKNNVYTCRHTRCNLGLCLSGNRECKED 592
Qy 421 CSDGDEKDCDCCGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480
Db 593 CSDGDEKDCDCCGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652
Qy 481 SAACVYIDRGFRYSPTQWTAFLGHDGQSRSAPOVQERRLKRIISHPPFNDFTPYDI 540
Db 653 SAACVYIDRGFRYSPTQWTAFLGHDGQSRSAPOVQERRLKRIISHPPFNDFTPYDI 712
Qy 541 ALLELEKPAEYSSMWRTICLPDASHVPPAKALWYMGHTQYGGTALILQGEIRVIN 600
Db 713 ALLELEKPAEYSSMWRTICLPDASHVPPAKALWYMGHTQYGGTALILQGEIRVIN 772
Qy 601 QTTCEMLLPQOITPRMNCVFLSGVDS CGDSGSGPLSSVEADGRIFQAGVSMGDSCAQ 660
Db 773 QTTCEMLLPQOITPRMNCVFLSGVDS CGDSGSGPLSSVEADGRIFQAGVSMGDSCAQ 832
Qy 661 RNKRGVYTRLPFRDWIKENTGV 683
Db 833 RNKRGVYTRLPFRDWIKENTGV 855
RESULT 6
US-10-072-012-354
Sequence 354, Application US/10072012
Publication No. US2004003493A1
GENERAL INFORMATION:
APPLICANT: Tchernev, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zerhusen, Bryan
APPLICANT: Patuturajan, Meera
APPLICANT: Shimkets, Richard
APPLICANT: Li, Li
APPLICANT: Gangoli, Beha
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie E.
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimyr Y.
APPLICANT: Colman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Putrak, Katarzyna
APPLICANT: Groese, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072.012
PRIOR FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
```

PRIOR APPLICATION NUMBER: 60/265,514  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,517  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,412  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,395  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/266,406  
PRIOR FILING DATE: 2001-02-02  
PRIOR APPLICATION NUMBER: 60/266,767  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: 60/267,057  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/266,975  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/267,459  
PRIOR FILING DATE: 2001-02-08  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1391  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 354  
LENGTH: 855  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-072-012-354

Query Match 99.1%; Score 3756; DB 15; Length 855;

Best Local Similarity 99.3%; Pred. No. 4.5e-275; Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 MAEERVVMLPPRARSLKSFVVTSVVAPPTDSKTVOQTQDSCSGFGLHARGVELMRFTTTPG 60  
173 MAEERVVMLPPRARSLKSFVVTSVVAPPTDSKTVOQTQDSCSGFGLHARGVELMRFTTTPG 232  
61 FPPSPYPAHARCOMALRGDADSVLSLTFRRSPDLASCBERSDLYVTYNTISPMEPALVQ 120  
233 FPPSPYPAHARCOMALRGDADSVLSLTFRRSPDLASCBERSDLYVTYNTISPMEPALVQ 292  
121 LCGTYPSSYMLTFHSSQNVLLTLITNTERRHQFEATFQPRMSCGRLKAQGTN 180  
293 LCGTYPSSYMLTFHSSQNVLLTLITNTERRHQFEATFQPRMSCGRLKAQGTN 352  
181 SPYYPGHYPNIDCTWNIIEVNNQHVVRPKFFYLEPRRACGTCPRDYVEINGEKYCGE 240  
353 SPYYPGHYPNIDCTWNIIEVNNQHVVRPKFFYLEPRRACGTCPRDYVEINGEKYCGE 412  
241 RSQFVVTNSNKTIVRHSQSTYTDGFLAEVLSYDSDCPCQFTCTGRCIRKELRCD 300  
413 RSQFVVTNSNKTIVRHSQSTYTDGFLAEVLSYDSDCPCQFTCTGRCIRKELRCD 472  
301 GMADCTHSDLELNCSCDAGHQTCKNFKCKPLFWVCDVNDGNSDEGSCSPACTQFRC 360  
473 GMADCTHSDLELNCSCDAGHQTCKNFKCKPLFWVCDVNDGNSDEGSCSPACTQFRC 532  
361 SNKGKLSKSQCKGKDCGDSDEASCPRXNVVYCTKHTYRCNLGLCLSKGNBECCKED 420  
533 SNKGKLSKSQCKGKDCGDSDEASCPRXNVVYCTKHTYRCNLGLCLSKGNBECCKED 592  
421 CSNGSDEKDDCGIRSTRROARVVGTDADGEMPMOVSJLHAGGHCICASLISPMWLV 480  
593 CSNGSDEKDDCGIRSTRROARVVGTDADGEMPMOVSJLHAGGHCICASLISPMWLV 652  
481 SAHACYIDDRGFRTDPTQWTAFLGLHDOSORSAPGQOEKRLKRIISHPFNDFTPDYDI 540  
653 SAHACYIDDRGFRTDPTQWTAFLGLHDOSORSAPGQOEKRLKRIISHPFNDFTPDYDI 712  
541 ALLELEKPAEYSSMWREICLPDASHVPACKAIWVTGHTOYGGTGALLIQRKEIRVIN 600  
713 ALLELEKPAEYSSMWREICLPDASHVPACKAIWVTGHTOYGGTGALLIQRKEIRVIN 772  
601 QTTCCENLLPQOITPRMMCVGFLSGVDSCGDSGGPISSEADGRITQAGVWSGDCCAQ 660

DB 773 QTTCCENLLPQOITPRMMCVGFLSGVDSCGDSGGPISSEADGRITQAGVWSGDCCAQ 832  
QY 661 RNKPGVYTRLPLEPRDWIKENTGV 683  
DB 833 RNKPGVYTRLPLEPRDWIKENTGV 855

## RESULT 7

US-10-072-012-420  
Sequence 420, Application US/10072012  
Publication No. US2004003493A1

## GENERAL INFORMATION:

APPLICANT: Tchernev, Velizar  
APPLICANT: Spytek, Kimberly  
APPLICANT: Zerhusen, Bryan  
APPLICANT: Patursajan, Meera  
APPLICANT: Shimkets, Richard  
APPLICANT: Li, Li  
APPLICANT: Gangolli, Bsha  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Anderson, David W.  
APPLICANT: Rastelli, Luca  
APPLICANT: Miller, Charles E.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Taupier Jr, Raymond J.  
APPLICANT: Gusev, Vladimir Y.  
APPLICANT: Colman, Steven D.  
APPLICANT: Wolenc, Adam R.  
APPLICANT: Pena, Carol E. A.  
APPLICANT: Furtak, Katarzyna  
APPLICANT: Grosse, William M.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Lepley, Denise M.  
APPLICANT: Rieger, Daniel K.  
APPLICANT: Burgess, Catherine E.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-258  
CURRENT APPLICATION NUMBER: US/10/072,012  
CURRENT FILING DATE: 2002-01-31  
PRIOR APPLICATION NUMBER: 60/265,102  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: 60/265,514  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,517  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,412  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,395  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/266,406  
PRIOR FILING DATE: 2001-02-02  
PRIOR APPLICATION NUMBER: 60/266,767  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: 60/267,057  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/266,975  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/267,459  
PRIOR FILING DATE: 2001-02-08  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1391  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 420  
LENGTH: 855  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-072-012-420

Query Match 99.1%; Score 3756; DB 15; Length 855;

Best Local Similarity 99.3%; Pred. No. 4.5e-275; Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 MAEERVVMLPPRARSLKSFVVTSVVAPPTDSKTVOQTQDSCSGFGLHARGVELMRFTTTPG 60

```

Db 173 MAERVVMLPPRARSLSFVVTSVAPFTDSKTQVORTDNCSCFGLHARGVELMRFTTPG 232
Qy 61 PPDSPPYPAHARCCQALRGDADSVLSTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 120
Db 233 PPDSPPYPAHARCCQALRGDADSVLSTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 292
Qy 121 LCGTTPPSYNLTFFHSSQNVLLITLITNTERRHGFEATFQLPMSSCGGLRKAOGTFN 180
Db 293 LCGTTPPSYNLTFFHSSQNVLLITLITNTERRHGFEATFQLPMSSCGGLRKAOGTFN 352
Qy 181 SPYYPGHYPPNIDCTWNIIEVNNQHVAKRFKFFYLEPRRACGTCPCDYOYVEINKEKXCGE 240
Db 353 SPYYPGHYPPNIDCTWNIIEVNNQHVAKRFKFFYLEPRRACGTCPCDYOYVEINKEKXCGE 412
Qy 241 RSQFVTVTSNKNKTVTRFHSQSYTDGTGFLAEYLSYDSSDPCPGQFTCRTRGCRIRKELRCD 300
Db 413 RSQFVTVTSNKNKTVTRFHSQSYTDGTGFLAEYLSYDSSDPCPGQFTCRTRGCRIRKELRCD 472
Qy 301 GMADCTHSDBLNCSGAGHOFCTCKNFKCPLFWVCDSVNDCCGNSDBEGSCSPAOQTFRC 360
Db 473 GMADCTHSDBLNCSGAGHOFCTCKNFKCPLFWVCDSVNDCCGNSDBEGSCSPAOQTFRC 532
Qy 361 SNGKCLSKSQOQCNKGKDCGSDGDEASCPKVVNVYCTKHTYRCLNGLCLSKGNPECDGKED 420
Db 533 SNGKCLSKSQOQCNKGKDCGSDGDEASCPKVVNVYCTKHTYRCLNGLCLSKGNPECDGKED 592
Qy 421 CSDGSDERKDCDGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSDGSDERKDCDGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAACHYIDDRGFRSDPTQWTAFLGLHDQSORSAFGVERLKRILISHPFNDFTFYDI 540
Db 653 SAACHYIDDRGFRSDPTQWTAFLGLHDQSORSAFGVERLKRILISHPFNDFTFYDI 712
Qy 541 ALLELEKPAEYXSNVPRICLPDASHVPAGKAIWYTGSHGYGAGTALLIQKGEIRVYN 600
Db 713 ALLELEKPAEYXSNVPRICLPDASHVPAGKAIWYTGSHGYGAGTALLIQKGEIRVYN 772
Qy 601 QTTCEMLLPQGITPRMNCVGLSGGVDSGQSDGGLSVBADRIIFQAGVAVWGDGCAQ 660
Db 773 QTTCEMLLPQGITPRMNCVGLSGGVDSGQSDGGLSVBADRIIFQAGVAVWGDGCAQ 832
Qy 661 RNKRGVYTRLPFLFRDWIKENTGV 683
Db 833 RNKRGVYTRLPFLFRDWIKENTGV 855

```

```

APPLICANT: Shenoy, Suresh G
APPLICANT: Anderson, David W
APPLICANT: Padigaru, Muralidhara
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Miller, Charles E
APPLICANT: Eileen, Andrew J
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-235
CURRENT APPLICATION NUMBER: US/10/037,417
CURRENT FILING DATE: 2002-09-20
PRIOR APPLICATION NUMBER: 60/260,018
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,360
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/272,411
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/272,817
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/291,186
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: 60/303,231
PRIOR FILING DATE: 2001-07-05
PRIOR APPLICATION NUMBER: 60/305,060
PRIOR FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 60/318,405
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/318,700
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 227
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 132
LENGTH: 855
TYPE: PRT
ORGANISM: Homo sapiens
US-10-037-417-132

Query Match 99.1%; Score 3756; DB 15; Length 855;
Best Local Similarity 99.3%; Pred. No. 4.5e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRARSLSFVVTSVAPFTDSKTQVORTDNCSCFGLHARGVELMRFTTPG 60
Db 173 MAERVVMLPPRARSLSFVVTSVAPFTDSKTQVORTDNCSCFGLHARGVELMRFTTPG 232
Qy 61 PPDSPPYPAHARCCQALRGDADSVLSTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 120
Db 233 PPDSPPYPAHARCCQALRGDADSVLSTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 292
Qy 121 LCGTTPPSYNLTFFHSSQNVLLITLITNTERRHGFEATFQLPMSSCGGLRKAOGTFN 180
Db 293 LCGTTPPSYNLTFFHSSQNVLLITLITNTERRHGFEATFQLPMSSCGGLRKAOGTFN 352
Qy 181 SPYYPGHYPPNIDCTWNIIEVNNQHVAKRFKFFYLEPRRACGTCPCDYOYVEINKEKXCGE 240
Db 353 SPYYPGHYPPNIDCTWNIIEVNNQHVAKRFKFFYLEPRRACGTCPCDYOYVEINKEKXCGE 412
Qy 241 RSQFVTVTSNKNKTVTRFHSQSYTDGTGFLAEYLSYDSSDPCPGQFTCRTRGCRIRKELRCD 300
Db 413 RSQFVTVTSNKNKTVTRFHSQSYTDGTGFLAEYLSYDSSDPCPGQFTCRTRGCRIRKELRCD 472
Qy 301 GMADCTHSDBLNCSGAGHOFCTCKNFKCPLFWVCDSVNDCCGNSDBEGSCSPAOQTFRC 360
Db 473 GMADCTHSDBLNCSGAGHOFCTCKNFKCPLFWVCDSVNDCCGNSDBEGSCSPAOQTFRC 532
Qy 361 SNGKCLSKSQOQCNKGKDCGSDGDEASCPKVVNVYCTKHTYRCLNGLCLSKGNPECDGKED 420
Db 533 SNGKCLSKSQOQCNKGKDCGSDGDEASCPKVVNVYCTKHTYRCLNGLCLSKGNPECDGKED 592
Qy 421 CSDGSDERKDCDGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSDGSDERKDCDGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAACHYIDDRGFRSDPTQWTAFLGLHDQSORSAFGVERLKRILISHPFNDFTFYDI 540

```

```
|||||
Db 653 SAACHYIDDDGFRYSPTQMTWTFGLGHDQSQRSAFGVQERRLKRILISHPFNFDPDYDI 712
Qy 541 ALLELEKPAEYSSMWPRICLPDASHVFPAGKAIWVTGMGHTOYGGTALILQGEIRVIN 600
Db 713 ALLELEKPAEYSSMWPRICLPDASHVFPAGKAIWVTGMGHTOYGGTALILQGEIRVIN 772
Qy 601 QTTCEMLPQOITPRMNCVGFSLGSGVDSCGDSGGLSVYADGRIFQAGVSWGDCAQ 660
Db 773 QTTCEMLPQOITPRMNCVGFSLGSGVDSCGDSGGLSVYADGRIFQAGVSWGDCAQ 832
Qy 661 RNKRGVYTRLPFRDWIKENTGV 683
Db 833 RNKRGVYTRLPFRDWIKENTGV 855

RESULT 9
US-10-072-012-44
; Sequence 44, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zehusen, Bryan
; APPLICANT: Patnubajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Esba
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Putak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgees, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 757
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-44
Query Match 99.1%; Score 3754; DB 15; Length 757;
Best Local Similarity 99.3%; Pred. No. 5.5e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAEERVVMLPPRRARSLKSFVYTSVVAFPPTDSKTQRTODNSCSFGHARGVBLMRFTTPG 60
Db 75 MAEERVVMLPPRRARSLKSFVYTSVVAFPPTDSKTQRTODNSCSFGHARGVBLMRFTTPG 134
Qy 61 FPDSPYPAHARCCQWALRGDASVLSITFRSFDLASCDERGSGLVYVNTLSPMEPHALVQ 120
Db 135 FPDSPYPAHARCCQWALRGDASVLSITFRSFDLASCDERGSGLVYVNTLSPMEPHALVQ 194
Qy 121 LCGTPEPSVNLTFHSSQNVLLITLINTBRHHPGFATFQOLPRMSSCGRLRKAQGTEN 180
Db 195 LCGTPEPSVNLTFHSSQNVLLITLINTBRHHPGFATFQOLPRMSSCGRLRKAQGTEN 254
Qy 181 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFFYLLEPRRACGTCRKYDYEINGEKYCGE 240
Db 255 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFFYLLEPGVPATCRKYDYEINGEKYCGE 314
Qy 241 RSQFVVTSSNKKITVRPHSDQSYDTTGFLAEVLSYDSSDPCQOFTCRGCRIRKRLCD 300
Db 315 RSQFVVTSSNKKITVRPHSDQSYDTTGFLAEVLSYDSSDPCQOFTCRGCRIRKRLCD 374
Qy 301 GWADCTDHSDELNCSGDAGHOFCKRKKPKPLFWVCDVNDGSDNDEGSCGPAQTFRC 360
Db 375 GWADCTDHSDELNCSGDAGHOFCKRKKPKPLFWVCDVNDGSDNDEGSCGPAQTFRC 434
Qy 361 SNGKCLSKSQQCKGKDCGDSDEASCPRKVVVTCRKHYYRCINGLCLSKGNECDGKED 420
Db 435 SNGKCLSKSQQCKGKDCGDSDEASCPRKVVVTCRKHYYRCINGLCLSKGNECDGKED 494
Qy 421 CSDGSEKDCDGLRFTQARVVGTDADDEGEMPVQVSLHALGQGHICGASLISPMVLV 480
Db 495 CSDGSEKDCDGLRFTQARVVGTDADDEGEMPVQVSLHALGQGHICGASLISPMVLV 554
Qy 481 SAACHYIDDDGFRYSPTQMTWTFGLGHDQSQRSAFGVQERRLKRILISHPFNFDPDYDI 540
Db 555 SAACHYIDDDGFRYSPTQMTWTFGLGHDQSQRSAFGVQERRLKRILISHPFNFDPDYDI 614
Qy 541 ALLELEKPAEYSSMWPRICLPDASHVFPAGKAIWVTGMGHTOYGGTALILQGEIRVIN 600
Db 615 ALLELEKPAEYSSMWPRICLPDASHVFPAGKAIWVTGMGHTOYGGTALILQGEIRVIN 674
Qy 601 QTTCEMLPQOITPRMNCVGFSLGSGVDSCGDSGGLSVYADGRIFQAGVSWGDCAQ 660
Db 675 QTTCEMLPQOITPRMNCVGFSLGSGVDSCGDSGGLSVYADGRIFQAGVSWGDCAQ 734
Qy 661 RNKRGVYTRLPFRDWIKENTGV 683
Db 735 RNKRGVYTRLPFRDWIKENTGV 757

RESULT 10
US-09-776-191-2
; Sequence 2, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiumn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
```



PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 60/234,840  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 60/179,982  
PRIOR FILING DATE: 2000-02-03  
PRIOR APPLICATION NUMBER: 60/183,542  
PRIOR FILING DATE: 2000-02-18  
PRIOR APPLICATION NUMBER: 09/657,968  
PRIOR FILING DATE: 2000-02-08  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 855  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-09-776-191-2

Query Match 99.1%; Score 3754; DB 10; Length 855;  
Best Local Similarity 99.3%; Pred. No. 6.4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVYMLPPRARSLSFVVTSVVAFPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 60  
DB 173 MAERVYMLPPRARSLSFVVTSVVAFPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 232  
QY 61 PPDSPPAHARCCQWALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120  
DB 233 PPDSPPAHARCCQWALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292  
QY 121 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPMSGCCGLRKAQGTEN 180  
DB 293 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPMSGCCGLRKAQGTEN 352  
QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLEBRACGTCPKDYVEINGEKYCGE 240  
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLEBRACGTCPKDYVEINGEKYCGE 412  
QY 241 RSQGVVTSNKNKTVRPHSDQSYTDGTGLAEVLSYDSDPCPGQFTCTGRCKIRKELRCD 300  
DB 413 RSQGVVTSNKNKTVRPHSDQSYTDGTGLAEVLSYDSDPCPGQFTCTGRCKIRKELRCD 472  
QY 301 GMADCTHSDLELNCSCDAGHOFCKNFKCKPLFWVCDSDVNDGDNDEQSCCPAQTFRC 360  
DB 473 GMADCTHSDLELNCSCDAGHOFCKNFKCKPLFWVCDSDVNDGDNDEQSCCPAQTFRC 532  
QY 361 SNGKCLSKSQCCNGKDCGDSDEASCPKXNVVYCTGHTYRCLNGLCLSKGNPECDGKED 420  
DB 533 SNGKCLSKSQCCNGKDCGDSDEASCPKXNVVYCTGHTYRCLNGLCLSKGNPECDGKED 592  
QY 421 CSDGSDKDCDCGRLRSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480  
DB 593 CSDGSDKDCDCGRLRSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652  
QY 481 SAACHYIDDBGFRYSDPTQWTAFLGLHDQSORAPGVQERLRKRIISHPEFNDFTFYDI 540  
DB 653 SAACHYIDDBGFRYSDPTQWTAFLGLHDQSORAPGVQERLRKRIISHPEFNDFTFYDI 712  
QY 541 ALLELEKPAEYSSWVRPCLPDAASHVPAGKAIWVTGMGHTQYGGTGLILQKGEIRVIN 600  
DB 713 ALLELEKPAEYSSWVRPCLPDAASHVPAGKAIWVTGMGHTQYGGTGLILQKGEIRVIN 772  
QY 601 QTTCEMLLPQOITPRMNCVFLSGGVDSCGDSCGGLSVLEADGRIFQAGVVSWGDCQAQ 660  
DB 773 QTTCEMLLPQOITPRMNCVFLSGGVDSCGDSCGGLSVLEADGRIFQAGVVSWGDCQAQ 832  
QY 661 RNKPGVYTRLPPLFRDWIKENTGV 683  
DB 833 RNKPGVYTRLPPLFRDWIKENTGV 855

RESULT 11  
US-10-099-700A-2  
Sequence 2, Application US/10099700A

Publication No. US20030008372A1  
GENERAL INFORMATION:  
APPLICANT: Edwin L. Madison  
APPLICANT: Edgar O. Ong  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7,  
FILE REFERENCE: 24745-1613  
CURRENT APPLICATION NUMBER: US/10/099,700A  
CURRENT FILING DATE: 2002-05-24  
PRIOR APPLICATION NUMBER: 60/275,592  
PRIOR FILING DATE: 2001-03-13  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 855  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-099-700A-2

Query Match 99.1%; Score 3754; DB 14; Length 855;  
Best Local Similarity 99.3%; Pred. No. 6.4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVYMLPPRARSLSFVVTSVVAFPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 60  
DB 173 MAERVYMLPPRARSLSFVVTSVVAFPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 232  
QY 61 PPDSPPAHARCCQWALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120  
DB 233 PPDSPPAHARCCQWALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292  
QY 121 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPMSGCCGLRKAQGTEN 180  
DB 293 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPMSGCCGLRKAQGTEN 352  
QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLEBRACGTCPKDYVEINGEKYCGE 240  
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLEBRACGTCPKDYVEINGEKYCGE 412  
QY 241 RSQGVVTSNKNKTVRPHSDQSYTDGTGLAEVLSYDSDPCPGQFTCTGRCKIRKELRCD 300  
DB 413 RSQGVVTSNKNKTVRPHSDQSYTDGTGLAEVLSYDSDPCPGQFTCTGRCKIRKELRCD 472  
QY 301 GMADCTHSDLELNCSCDAGHOFCKNFKCKPLFWVCDSDVNDGDNDEQSCCPAQTFRC 360  
DB 473 GMADCTHSDLELNCSCDAGHOFCKNFKCKPLFWVCDSDVNDGDNDEQSCCPAQTFRC 532  
QY 361 SNGKCLSKSQCCNGKDCGDSDEASCPKXNVVYCTGHTYRCLNGLCLSKGNPECDGKED 420  
DB 533 SNGKCLSKSQCCNGKDCGDSDEASCPKXNVVYCTGHTYRCLNGLCLSKGNPECDGKED 592  
QY 421 CSDGSDKDCDCGRLRSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480  
DB 593 CSDGSDKDCDCGRLRSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652  
QY 481 SAACHYIDDBGFRYSDPTQWTAFLGLHDQSORAPGVQERLRKRIISHPEFNDFTFYDI 540  
DB 653 SAACHYIDDBGFRYSDPTQWTAFLGLHDQSORAPGVQERLRKRIISHPEFNDFTFYDI 712  
QY 541 ALLELEKPAEYSSWVRPCLPDAASHVPAGKAIWVTGMGHTQYGGTGLILQKGEIRVIN 600  
DB 713 ALLELEKPAEYSSWVRPCLPDAASHVPAGKAIWVTGMGHTQYGGTGLILQKGEIRVIN 772  
QY 601 QTTCEMLLPQOITPRMNCVFLSGGVDSCGDSCGGLSVLEADGRIFQAGVVSWGDCQAQ 660  
DB 773 QTTCEMLLPQOITPRMNCVFLSGGVDSCGDSCGGLSVLEADGRIFQAGVVSWGDCQAQ 832  
QY 661 RNKPGVYTRLPPLFRDWIKENTGV 683  
DB 833 RNKPGVYTRLPPLFRDWIKENTGV 855

RESULT 12

US-10-190-030B-2  
; Sequence 2, Application US/10190030B  
; Publication No. US20030134298A1  
; GENERAL INFORMATION:  
; APPLICANT: Madison, Edwin  
; APPLICANT: Ong, Edgar  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A  
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND  
; TITLE OF INVENTION: METHODS BASED THEREON  
; FILE REFERENCE: 24745-1618  
; CURRENT APPLICATION NUMBER: US/10/190, 030B  
; CURRENT FILING DATE: 2002-07-03  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRF  
; ORGANISM: Homo Sapien  
US-10-190-030B-2

Query Match 99.1%; Score 3754; DB 14; Length 855;

Best Local Similarity 99.3%; Pred. No. 6.4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRARSLKSPVVTSVVAFPTDSKTVOQTODNSCSFGIHAAGVELMRFTTGG 60  
Db 173 MAERVVMLPPRARSLKSPVVTSVVAFPTDSKTVOQTODNSCSFGIHAAGVELMRFTTGG 232  
Qy 61 PPDSPPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120  
Db 233 PPDSPPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292  
Qy 121 LCGTTPSYNLTFRSSONVLLITLITNTERRHGFEATFPOLPRMSSCGRLKAQGTEN 180  
Db 293 LCGTTPSYNLTFRSSONVLLITLITNTERRHGFEATFPOLPRMSSCGRLKAQGTEN 352  
Qy 181 SPYPGHYPPNIDCTNIEVPNNQHVKRFKFFYLLEPRRACGTCPRDYVEINGEKYCGE 240  
Db 353 SPYPGHYPPNIDCTNIEVPNNQHVKRFKFFYLLEPRRACGTCPRDYVEINGEKYCGE 412  
Qy 241 RSQFVVTSSNKTITVRFHSDQSYTDGFLAEVLSYDSDPCPGQFCTRGRCIRKELRCD 300  
Db 413 RSQFVVTSSNKTITVRFHSDQSYTDGFLAEVLSYDSDPCPGQFCTRGRCIRKELRCD 472  
Qy 301 GMAADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDGDNDEGSCGCPAQTFRC 360  
Db 473 GMAADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDGDNDEGSCGCPAQTFRC 532  
Qy 361 SNGKCLSKSQOQCNKGKDCGDSDEASCPRKNVVTCTKHTYRCINGLCLSKGNPECDGKED 420  
Db 533 SNGKCLSKSQOQCNKGKDCGDSDEASCPRKNVVTCTKHTYRCINGLCLSKGNPECDGKED 592  
Qy 421 CSQDSBKDCDCLRSFTTRQARVVGSTADGEGMPQVSIHALGOGHICGASLISPMWLV 480  
Db 593 CSQDSBKDCDCLRSFTTRQARVVGSTADGEGMPQVSIHALGOGHICGASLISPMWLV 652  
Qy 481 SAACHYIDRGFRYSPTQWTAFLGLHDOSQASAPGOERRLKRIISHPFNDFTFDYDI 540  
Db 653 SAACHYIDRGFRYSPTQWTAFLGLHDOSQASAPGOERRLKRIISHPFNDFTFDYDI 712  
Qy 541 ALLELEKPAEYSSMWRPICLPDASHVFPAGKAIWVTGWHGTQYGGTALILQKEIRVIN 600  
Db 713 ALLELEKPAEYSSMWRPICLPDASHVFPAGKAIWVTGWHGTQYGGTALILQKEIRVIN 772  
Qy 601 QTTCEMLLPQOITPRMMCVGFLSGVDSGCGSDGSPISVEADGRIFOAGVSWGDCQAQ 660  
Db 773 QTTCEMLLPQOITPRMMCVGFLSGVDSGCGSDGSPISVEADGRIFOAGVSWGDCQAQ 832  
Qy 661 RNKPGVYTRLLPLFRDWIKENTGV 683  
Db 833 RNKPGVYTRLLPLFRDWIKENTGV 855

RESULT 13  
US-10-302-840A-2  
; Sequence 2, Application US/10302840A  
; Publication No. US20030134794A1  
; GENERAL INFORMATION:  
; APPLICANT: Madison, Edwin L.  
; APPLICANT: Ong, Edgar O.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCODED  
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON  
; FILE REFERENCE: 24745-1622  
; CURRENT APPLICATION NUMBER: US/10/302, 840A  
; CURRENT FILING DATE: 2003-01-24  
; PRIOR APPLICATION NUMBER: 60/332, 015  
; PRIOR FILING DATE: 2001-11-20  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRF  
; ORGANISM: Homo Sapien  
US-10-302-840A-2

Query Match 99.1%; Score 3754; DB 14; Length 855;

Best Local Similarity 99.3%; Pred. No. 6.4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRARSLKSPVVTSVVAFPTDSKTVOQTODNSCSFGIHAAGVELMRFTTGG 60  
Db 173 MAERVVMLPPRARSLKSPVVTSVVAFPTDSKTVOQTODNSCSFGIHAAGVELMRFTTGG 232  
Qy 61 PPDSPPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120  
Db 233 PPDSPPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292  
Qy 121 LCGTTPSYNLTFRSSONVLLITLITNTERRHGFEATFPOLPRMSSCGRLKAQGTEN 180  
Db 293 LCGTTPSYNLTFRSSONVLLITLITNTERRHGFEATFPOLPRMSSCGRLKAQGTEN 352  
Qy 181 SPYPGHYPPNIDCTNIEVPNNQHVKRFKFFYLLEPRRACGTCPRDYVEINGEKYCGE 240  
Db 353 SPYPGHYPPNIDCTNIEVPNNQHVKRFKFFYLLEPRRACGTCPRDYVEINGEKYCGE 412  
Qy 241 RSQFVVTSSNKTITVRFHSDQSYTDGFLAEVLSYDSDPCPGQFCTRGRCIRKELRCD 300  
Db 413 RSQFVVTSSNKTITVRFHSDQSYTDGFLAEVLSYDSDPCPGQFCTRGRCIRKELRCD 472  
Qy 301 GMAADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDGDNDEGSCGCPAQTFRC 360  
Db 473 GMAADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDGDNDEGSCGCPAQTFRC 532  
Qy 361 SNGKCLSKSQOQCNKGKDCGDSDEASCPRKNVVTCTKHTYRCINGLCLSKGNPECDGKED 420  
Db 533 SNGKCLSKSQOQCNKGKDCGDSDEASCPRKNVVTCTKHTYRCINGLCLSKGNPECDGKED 592  
Qy 421 CSQDSBKDCDCLRSFTTRQARVVGSTADGEGMPQVSIHALGOGHICGASLISPMWLV 480  
Db 593 CSQDSBKDCDCLRSFTTRQARVVGSTADGEGMPQVSIHALGOGHICGASLISPMWLV 652  
Qy 481 SAACHYIDRGFRYSPTQWTAFLGLHDOSQASAPGOERRLKRIISHPFNDFTFDYDI 540  
Db 653 SAACHYIDRGFRYSPTQWTAFLGLHDOSQASAPGOERRLKRIISHPFNDFTFDYDI 712  
Qy 541 ALLELEKPAEYSSMWRPICLPDASHVFPAGKAIWVTGWHGTQYGGTALILQKEIRVIN 600  
Db 713 ALLELEKPAEYSSMWRPICLPDASHVFPAGKAIWVTGWHGTQYGGTALILQKEIRVIN 772  
Qy 601 QTTCEMLLPQOITPRMMCVGFLSGVDSGCGSDGSPISVEADGRIFOAGVSWGDCQAQ 660  
Db 773 QTTCEMLLPQOITPRMMCVGFLSGVDSGCGSDGSPISVEADGRIFOAGVSWGDCQAQ 832  
Qy 661 RNKPGVYTRLLPLFRDWIKENTGV 683  
Db 833 RNKPGVYTRLLPLFRDWIKENTGV 855

RESULT 14  
US-10-267-219-2  
; Sequence 2, Application US/10267219  
; Publication No. US20030143219a1  
; GENERAL INFORMATION:  
; APPLICANT: Madison, Edwin  
; APPLICANT: Madison, Edwin  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2  
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON  
; FILE REFERENCE: 24745-1621  
; CURRENT APPLICATION NUMBER: US/10/267,219  
; CURRENT FILING DATE: 2002-10-08  
; PRIOR APPLICATION NUMBER: 60/328,530  
; PRIOR FILING DATE: 09-OCT-2001  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-267-219-2

Query Match 99.1%; Score 3754; DB 14; Length 855;  
Best Local Similarity 99.3%; Pred. No. 6,4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRASLSKSFVVTSVVAFPTDSKTVQRTQDSCSFGHLARGVELMRFTTPG 60  
DB 173 MAERVVMLEPPRASLSKSFVVTSVVAFPTDSKTVQRTQDSCSFGHLARGVELMRFTTPG 232

QY 61 FPDSPYPAHARQWALRGDADSVSLTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 120  
DB 233 FPDSPYPAHARQWALRGDADSVSLTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 292

QY 121 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPKMSCGGLRKAQGTEN 180  
DB 293 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPKMSCGGLRKAQGTEN 352

QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVYKVFYLLERBRACGTCPKDYVEINGEKYCGE 240  
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVYKVFYLLERBRACGTCPKDYVEINGEKYCGE 412

QY 241 RSQFVVTNSNKTITVRPHSDOSTYDTGFLAEYLSYDSSDPCPGQFTCTGRCTRKELRCD 300  
DB 413 RSQFVVTNSNKTITVRPHSDOSTYDTGFLAEYLSYDSSDPCPGQFTCTGRCTRKELRCD 472

QY 301 GMAUCTHSDDELNCSCDAGHOFCTCKNFKCKPLFWVCDSDVNDGSDNDEGSCSPAQTFR 360  
DB 473 GMAUCTHSDDELNCSCDAGHOFCTCKNFKCKPLFWVCDSDVNDGSDNDEGSCSPAQTFR 532

QY 361 SNGKCLSKSQOQNGKDCGSDGSDENASCPKXNVVVTCTKTYRCLNGLCLSKGNPECDKED 420  
DB 533 SNGKCLSKSQOQNGKDCGSDGSDENASCPKXNVVVTCTKTYRCLNGLCLSKGNPECDKED 592

QY 421 CSDSDSDKDCDCCGLRSFTTRQARVVGCTADAGEPMQVSLHALGQGHICGASLISPMWLV 480  
DB 593 CSDSDSDKDCDCCGLRSFTTRQARVVGCTADAGEPMQVSLHALGQGHICGASLISPMWLV 652

QY 481 SAACHYIDDRGFRYSDFPTQMTAFGLHDQSORSAAGVQERLRKRIISHPFNDFTFDYDI 540  
DB 653 SAACHYIDDRGFRYSDFPTQMTAFGLHDQSORSAAGVQERLRKRIISHPFNDFTFDYDI 712

QY 541 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGAALLQKEIRIVIN 600  
DB 713 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGAALLQKEIRIVIN 772

QY 601 QTTCENLIPQOITPRMNCVGLSGGVNDCQDSSGGLSVADGRIFPAGVAVSWGDCAQ 660  
DB 773 QTTCENLIPQOITPRMNCVGLSGGVNDCQDSSGGLSVADGRIFPAGVAVSWGDCAQ 832

QY 661 RNKPGVYTRLPFLFRDWIKENTGV 683

DB 833 RNKPGVYTRLPFLFRDWIKENTGV 855

RESULT 15  
US-10-112-221A-2  
; Sequence 2, Application US/10112221A  
; Publication No. US20030166851A1  
; GENERAL INFORMATION:  
; APPLICANT: Madison, Edwin  
; APPLICANT: Madison, Edwin  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9,  
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON  
; FILE REFERENCE: 24745-1615  
; CURRENT APPLICATION NUMBER: US/10/112,221A  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/279,228  
; PRIOR FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/291,501  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-112-221A-2

Query Match 99.1%; Score 3754; DB 14; Length 855;  
Best Local Similarity 99.3%; Pred. No. 6,4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRASLSKSFVVTSVVAFPTDSKTVQRTQDSCSFGHLARGVELMRFTTPG 60  
DB 173 MAERVVMLEPPRASLSKSFVVTSVVAFPTDSKTVQRTQDSCSFGHLARGVELMRFTTPG 232

QY 61 FPDSPYPAHARQWALRGDADSVSLTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 120  
DB 233 FPDSPYPAHARQWALRGDADSVSLTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 292

QY 121 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPKMSCGGLRKAQGTEN 180  
DB 293 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPKMSCGGLRKAQGTEN 352

QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVYKVFYLLERBRACGTCPKDYVEINGEKYCGE 240  
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVYKVFYLLERBRACGTCPKDYVEINGEKYCGE 412

QY 241 RSQFVVTNSNKTITVRPHSDOSTYDTGFLAEYLSYDSSDPCPGQFTCTGRCTRKELRCD 300  
DB 413 RSQFVVTNSNKTITVRPHSDOSTYDTGFLAEYLSYDSSDPCPGQFTCTGRCTRKELRCD 472

QY 301 GMAUCTHSDDELNCSCDAGHOFCTCKNFKCKPLFWVCDSDVNDGSDNDEGSCSPAQTFR 360  
DB 473 GMAUCTHSDDELNCSCDAGHOFCTCKNFKCKPLFWVCDSDVNDGSDNDEGSCSPAQTFR 532

QY 361 SNGKCLSKSQOQNGKDCGSDGSDENASCPKXNVVVTCTKTYRCLNGLCLSKGNPECDKED 420  
DB 533 SNGKCLSKSQOQNGKDCGSDGSDENASCPKXNVVVTCTKTYRCLNGLCLSKGNPECDKED 592

QY 421 CSDSDSDKDCDCCGLRSFTTRQARVVGCTADAGEPMQVSLHALGQGHICGASLISPMWLV 480  
DB 593 CSDSDSDKDCDCCGLRSFTTRQARVVGCTADAGEPMQVSLHALGQGHICGASLISPMWLV 652

QY 481 SAACHYIDDRGFRYSDFPTQMTAFGLHDQSORSAAGVQERLRKRIISHPFNDFTFDYDI 540  
DB 653 SAACHYIDDRGFRYSDFPTQMTAFGLHDQSORSAAGVQERLRKRIISHPFNDFTFDYDI 712

QY 541 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGAALLQKEIRIVIN 600  
DB 713 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGAALLQKEIRIVIN 772

QY 601 QTTCENLIPQOITPRMNCVGLSGGVNDCQDSSGGLSVADGRIFPAGVAVSWGDCAQ 660

Db 773 QTTCEMLPQOITPRMNCVGLSGVDSCGDSGGPLSSVEADGRIFQAGVSVWGDCQAQ 832  
Qy 661 RNKGVYTRLPFRDWIKENTGV 683  
Db 833 RNKGVYTRLPFRDWIKENTGV 855

## RESULT 16

US-10-104-271-2  
; Sequence 2, Application US/10104271  
; Publication No. US20030181658A1  
; GENERAL INFORMATION:  
; APPLICANT: Madison, Edwin  
; APPLICANT: Jiumn-Chern Yeh  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVP14, THE ENCOI  
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON  
; FILE REFERENCE: 24745-1614  
; CURRENT APPLICATION NUMBER: US/10/104,271  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/278,166  
; PRIOR FILING DATE: 2001-03-22  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-104-271-2

Query Match 99.1%; Score 3754; DB 14; Length 855;  
Best Local Similarity 99.3%; Pred. No. 6,4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAEERVVMLPPRARSLKSFVTVSVVAFPTDSKTQVORTQDNCSCFGLHARGVELMRFTTGG 60  
Db 173 MAEERVVMLPPRARSLKSFVTVSVVAFPTDSKTQVORTQDNCSCFGLHARGVELMRFTTGG 232  
Qy 61 FPDSPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLTVYNTLSPMEPHALVQ 120  
Db 233 FPDSPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLTVYNTLSPMEPHALVQ 292  
Qy 121 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLKKAQGTEN 180  
Db 293 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLKKAQGTEN 352  
Qy 181 SPYYPGHYPENIDCTNMIIEVPNNQHVVRPKFFYLLEPRRACGTCRKYVEINGEKXCGE 240  
Db 353 SPYYPGHYPENIDCTNMIIEVPNNQHVVRPKFFYLLEPRRACGTCRKYVEINGEKXCGE 412  
Qy 241 RSQFVTVSNKLTIVRPHSDQSYTDGTGFLAEVLSYDSDPCPGQFCTGRCTRKEIRCD 300  
Db 413 RSQFVTVSNKLTIVRPHSDQSYTDGTGFLAEVLSYDSDPCPGQFCTGRCTRKEIRCD 472  
Qy 301 GMADCTDHSDELNCSCDAGHGFCTCKNFKCKPLFWVCDSVNDCGDSDGCGSCPAQTFRC 360  
Db 473 GMADCTDHSDELNCSCDAGHGFCTCKNFKCKPLFWVCDSVNDCGDSDGCGSCPAQTFRC 532  
Qy 361 SNGKCLSKSQOQCKGDKDCGDSDEASCPKXNVVTCYKHYRCLNGLCLSKGNPECDGKED 420  
Db 533 SNGKCLSKSQOQCKGDKDCGDSDEASCPKXNVVTCYKHYRCLNGLCLSKGNPECDGKED 592  
Qy 421 CSQDSDEKDCDGLRSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480  
Db 593 CSQDSDEKDCDGLRSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652  
Qy 481 SAHCHYIDDRGRYSPTQMTAFGLHDQSORSAAPGVQERRLKRIISHPFNDFTFDYDI 540  
Db 653 SAHCHYIDDRGRYSPTQMTAFGLHDQSORSAAPGVQERRLKRIISHPFNDFTFDYDI 712  
Qy 541 ALLELEKPAEYSSWVRICLPDASHVFPAGKAIWVTGWHGTQYGGTALLILOKGEIRVIN 600  
Db 713 ALLELEKPAEYSSWVRICLPDASHVFPAGKAIWVTGWHGTQYGGTALLILOKGEIRVIN 772

Qy 601 QTTCEMLPQOITPRMNCVGLSGVDSCGDSGGPLSSVEADGRIFQAGVSVWGDCQAQ 660  
Db 773 QTTCEMLPQOITPRMNCVGLSGVDSCGDSGGPLSSVEADGRIFQAGVSVWGDCQAQ 832  
Qy 661 RNKGVYTRLPFRDWIKENTGV 683  
Db 833 RNKGVYTRLPFRDWIKENTGV 855

## RESULT 17

US-10-147-211A-2  
; Sequence 2, Application US/10147211A  
; Publication No. US20030235900A1  
; GENERAL INFORMATION:  
; APPLICANT: Madison, Edwin  
; APPLICANT: Yeh, Jiumn-Chern  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1  
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON  
; FILE REFERENCE: 24745-1616  
; CURRENT APPLICATION NUMBER: US/10/147,211A  
; CURRENT FILING DATE: 2002-05-14  
; PRIOR APPLICATION NUMBER: 60/291,001  
; PRIOR FILING DATE: 2001-05-14  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-147-211A-2

Query Match 99.1%; Score 3754; DB 14; Length 855;  
Best Local Similarity 99.3%; Pred. No. 6,4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAEERVVMLPPRARSLKSFVTVSVVAFPTDSKTQVORTQDNCSCFGLHARGVELMRFTTGG 60  
Db 173 MAEERVVMLPPRARSLKSFVTVSVVAFPTDSKTQVORTQDNCSCFGLHARGVELMRFTTGG 232  
Qy 61 FPDSPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLTVYNTLSPMEPHALVQ 120  
Db 233 FPDSPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLTVYNTLSPMEPHALVQ 292  
Qy 121 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLKKAQGTEN 180  
Db 293 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLKKAQGTEN 352  
Qy 181 SPYYPGHYPENIDCTNMIIEVPNNQHVVRPKFFYLLEPRRACGTCRKYVEINGEKXCGE 240  
Db 353 SPYYPGHYPENIDCTNMIIEVPNNQHVVRPKFFYLLEPRRACGTCRKYVEINGEKXCGE 412  
Qy 241 RSQFVTVSNKLTIVRPHSDQSYTDGTGFLAEVLSYDSDPCPGQFCTGRCTRKEIRCD 300  
Db 413 RSQFVTVSNKLTIVRPHSDQSYTDGTGFLAEVLSYDSDPCPGQFCTGRCTRKEIRCD 472  
Qy 301 GMADCTDHSDELNCSCDAGHGFCTCKNFKCKPLFWVCDSVNDCGDSDGCGSCPAQTFRC 360  
Db 473 GMADCTDHSDELNCSCDAGHGFCTCKNFKCKPLFWVCDSVNDCGDSDGCGSCPAQTFRC 532  
Qy 361 SNGKCLSKSQOQCKGDKDCGDSDEASCPKXNVVTCYKHYRCLNGLCLSKGNPECDGKED 420  
Db 533 SNGKCLSKSQOQCKGDKDCGDSDEASCPKXNVVTCYKHYRCLNGLCLSKGNPECDGKED 592  
Qy 421 CSQDSDEKDCDGLRSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480  
Db 593 CSQDSDEKDCDGLRSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652  
Qy 481 SAHCHYIDDRGRYSPTQMTAFGLHDQSORSAAPGVQERRLKRIISHPFNDFTFDYDI 540  
Db 653 SAHCHYIDDRGRYSPTQMTAFGLHDQSORSAAPGVQERRLKRIISHPFNDFTFDYDI 712  
Qy 541 ALLELEKPAEYSSWVRICLPDASHVFPAGKAIWVTGWHGTQYGGTALLILOKGEIRVIN 600

Db 713 ALLELEKPAVSSWVRPCLPDPASHVFPAGKAIWVTGWHGTQYGTGALLQKEIRVIN 772  
Qy 601 QTTCEMLLPQOITPRMNCVGFSLSGVDSCQDGGPILSVADGRIFQAGVWSMGDCQAQ 660  
Db 773 QTTCEMLLPQOITPRMNCVGFSLSGVDSCQDGGPILSVADGRIFQAGVWSMGDCQAQ 832  
Qy 661 RNKGVYTRLPFRDWIKENTGV 683  
Db 833 RNKGVYTRLPFRDWIKENTGV 855

## RESULT 18

US-10-156-214A-2  
Sequence 2, Application US/10156214A  
Publication No. US20040001801A1  
GENERAL INFORMATION:  
APPLICANT: Edwin L. Madison  
APPLICANT: Joseph Edward Sempke  
APPLICANT: George P. Vlausk  
APPLICANT: Scott Jeffrey Kemp  
APPLICANT: Malarreddy Komandla  
APPLICANT: Daniel Vanna Slev  
TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic Use  
TITLE OF INVENTION: Thereof  
FILE REFERENCE: 24745-1611  
CURRENT APPLICATION NUMBER: US/10/156,214A  
CURRENT FILING DATE: 2002-05-23  
NUMBER OF SEQ ID NOS: 611  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 855  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-156-214A-2

Query Match 99.1%; Score 3754; DB 15; Length 855;  
Best Local Similarity 99.3%; Pred. No. 6.4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRARSLKSFVVTSVVAFPTDSKTQVORTDNSCSFGLHARGVELMRFTTPG 60  
Db 173 MAERVVMLPPRARSLKSFVVTSVVAFPTDSKTQVORTDNSCSFGLHARGVELMRFTTPG 232  
Qy 61 PPDSPPYPAHARCCQALGADSVSLTFRSFDLASCBERSGDLVTVNTTSPMEPALVQ 120  
Db 233 PPDSPPYPAHARCCQALGADSVSLTFRSFDLASCBERSGDLVTVNTTSPMEPALVQ 292  
Qy 121 LCGTYPSPSYNLTFPSSQNVLLITLITNTERHPGFATFPQLPRMSSCGRLRAQGTFN 180  
Db 293 LCGTYPSPSYNLTFPSSQNVLLITLITNTERHPGFATFPQLPRMSSCGRLRAQGTFN 352  
Qy 181 SPYYPGHYPPNIDCTWNIIEVNNQHVVRFEFFYLLEPRACGTCPRDYVEINKEKTCGE 240  
Db 353 SPYYPGHYPPNIDCTWNIIEVNNQHVVRFEFFYLLEPRACGTCPRDYVEINKEKTCGE 412  
Qy 241 RSQGVTVTSNKNKIVTRHSDQSYTDGTFLAEYLSYSSDPGQFTGRTGKCIKEIRCD 300  
Db 413 RSQGVTVTSNKNKIVTRHSDQSYTDGTFLAEYLSYSSDPGQFTGRTGKCIKEIRCD 472  
Qy 301 GMADCTHSDLELNCSCDAGHOFCTCKNFKCKPLFWVCSVNDGNSDDEGSCPAQTFRC 360  
Db 473 GMADCTHSDLELNCSCDAGHOFCTCKNFKCKPLFWVCSVNDGNSDDEGSCPAQTFRC 532  
Qy 361 SNGKCLSKSQCCNGKDDCGSDDEASCPKVVVVTCTKHYRCLNGLCLSKNPECDKED 420  
Db 533 SNGKCLSKSQCCNGKDDCGSDDEASCPKVVVVTCTKHYRCLNGLCLSKNPECDKED 592  
Qy 421 CSDDSDKDDCCGGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480  
Db 593 CSDDSDKDDCCGGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652  
Qy 481 SAHHCYIDDRGFRSDPTQWTAFLGLHDGSRSAFGVQERLKRILSHPFNDFTFYDI 540

Db 653 SAHHCYIDDRGFRSDPTQWTAFLGLHDGSRSAFGVQERLKRILSHPFNDFTFYDI 712  
Qy 541 ALLELEKPAVSSWVRPCLPDPASHVFPAGKAIWVTGWHGTQYGTGALLQKEIRVIN 600  
Db 713 ALLELEKPAVSSWVRPCLPDPASHVFPAGKAIWVTGWHGTQYGTGALLQKEIRVIN 772  
Qy 601 QTTCEMLLPQOITPRMNCVGFSLSGVDSCQDGGPILSVADGRIFQAGVWSMGDCQAQ 660  
Db 773 QTTCEMLLPQOITPRMNCVGFSLSGVDSCQDGGPILSVADGRIFQAGVWSMGDCQAQ 832  
Qy 661 RNKGVYTRLPFRDWIKENTGV 683  
Db 833 RNKGVYTRLPFRDWIKENTGV 855

## RESULT 19

US-10-072-012-352  
Sequence 352, Application US/10072012  
Publication No. US2004003493A1  
GENERAL INFORMATION:  
APPLICANT: Tchenev, Velizar  
APPLICANT: Spytek, Kimberly  
APPLICANT: Zerhusen, Bryan  
APPLICANT: Patuturajan, Meera  
APPLICANT: Shinkets, Richard  
APPLICANT: Li, Li  
APPLICANT: Gangoli, Bsha  
APPLICANT: Padigaru, Muraidhara  
APPLICANT: Anderson, David W.  
APPLICANT: Rastelli, Luca  
APPLICANT: Miller, Charles E.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Taupier Jr, Raymond J.  
APPLICANT: Gusev, Vladimir Y.  
APPLICANT: Coleman, Steven D.  
APPLICANT: Wolenc, Adam R.  
APPLICANT: Pena, Carol E. A  
APPLICANT: Furtak, Katarzyna  
APPLICANT: Grose, William M.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Lepley, Denise M.  
APPLICANT: Rieger, Daniel K.  
APPLICANT: Burgess, Catherine E.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-258  
CURRENT APPLICATION NUMBER: US/10/072,012  
CURRENT FILING DATE: 2002-01-31  
PRIOR APPLICATION NUMBER: 60/265,102  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: 60/265,514  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,517  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,412  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,395  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/266,406  
PRIOR FILING DATE: 2001-02-02  
PRIOR APPLICATION NUMBER: 60/266,767  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: 60/267,057  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/266,975  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/267,459  
PRIOR FILING DATE: 2001-02-08  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1391  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 352  
LENGTH: 855

TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-072-012-352

Query Match 99.1%; Score 3754; DB 15; Length 855;

Best Local Similarity 99.3%; Pred. No. 6,4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 MAERVVMLEPPRARSLSKSPVTSVVAFPPTDSKTVOQTODNSCSFGLHARGVELMRFTTTPG 60
DB 173 MAERVVMLEPPRARSLSKSPVTSVVAFPPTDSKTVOQTODNSCSFGLHARGVELMRFTTTPG 232
QY 61 PPSPPYPAHARCQWALRGDADSVLSITFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQ 120
DB 233 PPSPPYPAHARCQWALRGDADSVLSITFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQ 292
QY 121 LCGTTPPSVNLTFHSSQNVLLITLITNTERRHHPGFATFPQLPRMSSCGRLRKAQGTN 180
DB 293 LCGTTPPSVNLTFHSSQNVLLITLITNTERRHHPGFATFPQLPRMSSCGRLRKAQGTN 352
QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFEFYLLPBRACGTCPPDYVEINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFEFYLLPBRACGTCPPDYVEINGEKYCGE 412
QY 241 RSQFVVTSSNKTITVRFHSDQSYTDGTGFLAEVLSYDSSDPCPGQFCRTGRCTRKELRCD 300
DB 413 RSQFVVTSSNKTITVRFHSDQSYTDGTGFLAEVLSYDSSDPCPGQFCRTGRCTRKELRCD 472
QY 301 GMDADCTHSDBELNCSCDAGHOFCKKFKPLFWVCDVNDGCGNSDBOGSCCPAQTFRC 360
DB 473 GMDADCTHSDBELNCSCDAGHOFCKKFKPLFWVCDVNDGCGNSDBOGSCCPAQTFRC 532
QY 361 SNGKCLSKSQOQCNKGKDCGSDGDEASCPKVVVYCTGHYTRCLNGLCLSKGNPECDKED 420
DB 533 SNGKCLSKSQOQCNKGKDCGSDGDEASCPKVVVYCTGHYTRCLNGLCLSKGNPECDKED 592
QY 421 CSDGSDKDCDCGIRSTRTQARVVGGTDADGEMPMQVSLHAGQGHICGASLISPMVLV 480
DB 593 CSDGSDKDCDCGIRSTRTQARVVGGTDADGEMPMQVSLHAGQGHICGASLISPMVLV 652
QY 481 SAHACVYDDGFRSDPTQMTAFGLHDOSORSAPOVERLRKRIISHPFNDPTPYDI 540
DB 653 SAHACVYDDGFRSDPTQMTAFGLHDOSORSAPOVERLRKRIISHPFNDPTPYDI 712
QY 541 ALLELEKPAEYSSWVRPICKLPDASHVPPAGKAIWVTMGHTQYGGTGALLQKEIRVIN 600
DB 713 ALLELEKPAEYSSWVRPICKLPDASHVPPAGKAIWVTMGHTQYGGTGALLQKEIRVIN 772
QY 601 QTTCENILLPOQITPRMCCVGFSLSGGVDSCOGDSGGPISVYADGRIFQAGVSWGDCAQ 660
DB 773 QTTCENILLPOQITPRMCCVGFSLSGGVDSCOGDSGGPISVYADGRIFQAGVSWGDCAQ 832
QY 661 RNKRGVYTRLLPLPDMWKENTGV 683
DB 833 RNKRGVYTRLLPLPDMWKENTGV 855

```

RESULT 20

US-10-072-012-411

Sequence 411, Application US/10072012

Publication No. US2004003493A1

GENERAL INFORMATION:

APPLICANT: Tchernyev, Velizar

APPLICANT: Spyttek, Kimberly

APPLICANT: Zeithusen, Bryan

APPLICANT: Paturajan, Meera

APPLICANT: Shukets, Richard

APPLICANT: Li, Li

APPLICANT: Gangoli, Esha

APPLICANT: Padigaru, Muralidhara

APPLICANT: Anderson, David W.

APPLICANT: Rastelli, Luca

APPLICANT: Miller, Charles E.

```

APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Colman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 411
LENGTH: 855
TYPE: PRT
ORGANISM: Homo sapiens
US-10-072-012-411

```

Query Match 99.1%; Score 3754; DB 15; Length 855;  
Best Local Similarity 99.3%; Pred. No. 6,4e-275;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 MAERVVMLEPPRARSLSKSPVTSVVAFPPTDSKTVOQTODNSCSFGLHARGVELMRFTTTPG 60
DB 173 MAERVVMLEPPRARSLSKSPVTSVVAFPPTDSKTVOQTODNSCSFGLHARGVELMRFTTTPG 232
QY 61 PPSPPYPAHARCQWALRGDADSVLSITFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQ 120
DB 233 PPSPPYPAHARCQWALRGDADSVLSITFRSFDLASCDERGSDLVTVYNTLSPMEPHALVQ 292
QY 121 LCGTTPPSVNLTFHSSQNVLLITLITNTERRHHPGFATFPQLPRMSSCGRLRKAQGTN 180
DB 293 LCGTTPPSVNLTFHSSQNVLLITLITNTERRHHPGFATFPQLPRMSSCGRLRKAQGTN 352
QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFEFYLLPBRACGTCPPDYVEINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFEFYLLPBRACGTCPPDYVEINGEKYCGE 412
QY 241 RSQFVVTSSNKTITVRFHSDQSYTDGTGFLAEVLSYDSSDPCPGQFCRTGRCTRKELRCD 300
DB 413 RSQFVVTSSNKTITVRFHSDQSYTDGTGFLAEVLSYDSSDPCPGQFCRTGRCTRKELRCD 472
QY 301 GMDADCTHSDBELNCSCDAGHOFCKKFKPLFWVCDVNDGCGNSDBOGSCCPAQTFRC 360
DB 473 GMDADCTHSDBELNCSCDAGHOFCKKFKPLFWVCDVNDGCGNSDBOGSCCPAQTFRC 532

```

```
OY 361 SNGKCLSKSQOQNGKDGSDGSDASCPKXNVVYCTKHTYRCLNGLCLSKGNPECDKED 420
|||
DB 533 SNGKCLSKSQOQNGKDGSDGSDASCPKXNVVYCTKHTYRCLNGLCLSKGNPECDKED 592
OY 421 CSDGSDKDCDCGRLSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480
CSDGSDKDCDCGRLSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652
DB 593 CSDGSDKDCDCGRLSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652
OY 481 SAACHYIDDRGFRSDPTQMTAFGLHDOSQASAPGVQERLKRILISHPFNDFTFYDI 540
543 SAACHYIDDRGFRSDPTQMTAFGLHDOSQASAPGVQERLKRILISHPFNDFTFYDI 712
OY 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGMGTGYGSGTALILQKEIRIVIN 600
|||
DB 713 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGMGTGYGSGTALILQKEIRIVIN 772
OY 601 QTTCEMLLPQOITPRMVCVFLSGGVDSGQDGGPLSVYADGRIFQAGVSWGDCQAQ 660
773 QTTCEMLLPQOITPRMVCVFLSGGVDSGQDGGPLSVYADGRIFQAGVSWGDCQAQ 832
DB 773 QTTCEMLLPQOITPRMVCVFLSGGVDSGQDGGPLSVYADGRIFQAGVSWGDCQAQ 832
OY 661 RNKRGVYTRLPFRDWIKENTGV 683
DB 833 RNKRGVYTRLPFRDWIKENTGV 855
```

```
RESULT 21
US-10-072-012-418
; Sequence 418, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Zephusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimr Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
```

```
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO: 418
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-418
```

```
Query Match 99.1%; Score 3754; DB 15; Length 855;
Best Local Similarity 99.3%; Pred. No. 6,4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 1 MAERVYMLPPRARSLSKSFVYTSVAPPTDSKTVQRTQDNSCSFGLHARGVELMRPTTPG 60
DB 173 MAERVYMLPPRARSLSKSFVYTSVAPPTDSKTVQRTQDNSCSFGLHARGVELMRPTTPG 232
OY 61 FPDSPYPAHARQWALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPEPHALVQ 120
DB 233 FPDSPYPAHARQWALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPEPHALVQ 292
OY 121 LCGTTPPSYNTLTFHSSQNVLLITLITTEBRRHGFETFPQLRMSCGRLRKAQSTFN 180
DB 293 LCGTTPPSYNTLTFHSSQNVLLITLITTEBRRHGFETFPQLRMSCGRLRKAQSTFN 352
OY 181 SPYYPGHYPNIDCTWNI EVPNNOHVVRFKFYLLEPRACGTCPPDYVEINKEKTCGE 240
DB 353 SPYYPGHYPNIDCTWNI EVPNNOHVVRFKFYLLEPRACGTCPPDYVEINKEKTCGE 412
OY 241 RSQFVYTSNSNKTITVRHSDQSYTDGFLAEYLSYDSSDPCQOFTCTGRCTRKELRCD 300
DB 413 RSQFVYTSNSNKTITVRHSDQSYTDGFLAEYLSYDSSDPCQOFTCTGRCTRKELRCD 472
OY 301 GMADCTDHSDELNCSCDAGHOFCKKFKCPFLWVCDSVNDCCDNDSDGSCDPAQTFRC 360
DB 473 GMADCTDHSDELNCSCDAGHOFCKKFKCPFLWVCDSVNDCCDNDSDGSCDPAQTFRC 532
OY 361 SNGKCLSKSQOQNGKDGSDGSDASCPKXNVVYCTKHTYRCLNGLCLSKGNPECDKED 420
DB 533 SNGKCLSKSQOQNGKDGSDGSDASCPKXNVVYCTKHTYRCLNGLCLSKGNPECDKED 592
OY 421 CSDGSDKDCDCGRLSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480
DB 593 CSDGSDKDCDCGRLSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652
OY 481 SAACHYIDDRGFRSDPTQMTAFGLHDOSQASAPGVQERLKRILISHPFNDFTFYDI 540
DB 543 SAACHYIDDRGFRSDPTQMTAFGLHDOSQASAPGVQERLKRILISHPFNDFTFYDI 712
OY 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGMGTGYGSGTALILQKEIRIVIN 600
DB 713 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGMGTGYGSGTALILQKEIRIVIN 772
OY 601 QTTCEMLLPQOITPRMVCVFLSGGVDSGQDGGPLSVYADGRIFQAGVSWGDCQAQ 660
DB 773 QTTCEMLLPQOITPRMVCVFLSGGVDSGQDGGPLSVYADGRIFQAGVSWGDCQAQ 832
OY 661 RNKRGVYTRLPFRDWIKENTGV 683
DB 833 RNKRGVYTRLPFRDWIKENTGV 855
```

```
RESULT 22
US-10-600-187-2
; Sequence 2, Application US/10600187
; Publication No. US20040086910A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
```



```
/ TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
/ FILE REFERENCE: D6064CIP/D
/ CURRENT APPLICATION NUMBER: US/10/600,187
/ CURRENT FILING DATE: 2003-06-20
/ PRIOR APPLICATION NUMBER: US/09/654,600A
/ PRIOR FILING DATE: 2000-09-01
/ PRIOR APPLICATION NUMBER: 09/421,213
/ 09/027,337
/ PRIOR FILING DATE: 1999-10-20
/ 1998-02-20
/ NUMBER OF SEQ ID NOS: 98
/ SEQ ID NO 2
/ LENGTH: 855
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: TADG-15
/ US-10-600-187-2
```

```
Query Match          99.1%; Score 3754; DB 15; Length 855;
Best Local Similarity 99.3%; Pred. No. 6.4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 MAEERVVMLPPRRASLSKSPVVTSVVAFPTDSKTIVQRTQDNSCSFGIHLARGVELMRFTTPG 60
DB 173 MAEERVVMLPPRRASLSKSPVVTSVVAFPTDSKTIVQRTQDNSCSFGIHLARGVELMRFTTPG 232
QY 61 FPDSPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYVNTLSPMEPHALVQ 120
DB 233 FPDSPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYVNTLSPMEPHALVQ 292
QY 121 LCGTTPPSYVNLTFHSSQNVLLITLINTERRHGFATFPQLPRMSSCGRLRKAQGTN 180
DB 293 LCGTTPPSYVNLTFHSSQNVLLITLINTERRHGFATFPQLPRMSSCGRLRKAQGTN 352
QY 181 SPYYPGHYPNIDCTNMIIEVNNQHVAKRFKFFYLLEPPRACGTCPKDYVEINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTNMIIEVNNQHVAKRFKFFYLLEPPRACGTCPKDYVEINGEKYCGE 412
QY 241 RSQFVVTSSNKKITVRFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCD 300
DB 413 RSQFVVTSSNKKITVRFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCD 472
QY 301 GMAADCTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDSVNDCGNSDEGSCSPAQTFRC 360
DB 473 GMAADCTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDSVNDCGNSDEGSCSPAQTFRC 532
QY 361 SNGKCLSKSQOQCKGDKDCGDSDEASCPRKVVVYCTKHYYRCINGLCLSKGNPECDGKED 420
DB 533 SNGKCLSKSQOQCKGDKDCGDSDEASCPRKVVVYCTKHYYRCINGLCLSKGNPECDGKED 592
QY 421 CSDGSDKDCDCCGLRSTFRQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480
DB 593 CSDGSDKDCDCCGLRSTFRQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652
QY 481 SAACHYIDRGRFYSPDTQMTAFGLHDOSQSAAPGVQERLRKRIISHPEFNDFTDYDI 540
DB 653 SAACHYIDRGRFYSPDTQMTAFGLHDOSQSAAPGVQERLRKRIISHPEFNDFTDYDI 712
QY 541 ALLELEKPAEYSSMWRPCLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 600
DB 713 ALLELEKPAEYSSMWRPCLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 772
QY 601 QTTCEMLPQOITPRMNCVGFISGVNDCGDSGGLSSVEADGRITQAGVSWGDCQAQ 660
DB 773 QTTCEMLPQOITPRMNCVGFISGVNDCGDSGGLSSVEADGRITQAGVSWGDCQAQ 832
QY 661 RNKRGVYTRILPLFRDWIKENTGV 683
DB 833 RNKRGVYTRILPLFRDWIKENTGV 855
```

```
RESULT 23
US-10-276-774-1798
/ Sequence 1798, Application US/10276774
/ Publication No. US20040053245A1
/ GENERAL INFORMATION:
/ APPLICANT: Hyseq, Inc.
/ APPLICANT: Tang, Y, Tom et al
/ TITLE OF INVENTION: No. US20040053245A1e1 Nucleic Acids and Polypeptides
/ FILE REFERENCE: 21272-030
/ CURRENT APPLICATION NUMBER: US/10/276,774
/ CURRENT FILING DATE: 2002-11-18
/ PRIOR APPLICATION NUMBER: 09/560,875
/ PRIOR FILING DATE: 2000-04-27
/ PRIOR APPLICATION NUMBER: 09/496,914
/ PRIOR FILING DATE: 2000-02-03
/ NUMBER OF SEQ ID NOS: 2700
/ SOFTWARE: Custom
/ SEQ ID NO 1798
/ LENGTH: 851
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-10-276-774-1798
```

```
Query Match          98.8%; Score 3745; DB 15; Length 851;
Best Local Similarity 99.0%; Pred. No. 3e-274;
Matches 676; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 MAEERVVMLPPRRASLSKSPVVTSVVAFPTDSKTIVQRTQDNSCSFGIHLARGVELMRFTTPG 60
DB 169 MAEERVVMLPPRRASLSKSPVVTSVVAFPTDSKTIVQRTQDNSCSFGIHLARGVELMRFTTPG 228
QY 61 FPDSPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYVNTLSPMEPHALVQ 120
DB 229 FPDSPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGHILVTYVNTLSPMEPHALVQ 288
QY 121 LCGTTPPSYVNLTFHSSQNVLLITLINTERRHGFATFPQLPRMSSCGRLRKAQGTN 180
DB 289 LCGTTPPSYVNLTFHSSQNVLLITLINTERRHGFATFPQLPRMSSCGRLRKAQGTN 348
QY 181 SPYYPGHYPNIDCTNMIIEVNNQHVAKRFKFFYLLEPPRACGTCPKDYVEINGEKYCGE 240
DB 349 SPYYPGHYPNIDCTNMIIEVNNQHVAKRFKFFYLLEPPRACGTCPKDYVEINGEKYCGE 408
QY 241 RSQFVVTSSNKKITVRFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCD 300
DB 409 RSQFVVTSSNKKITVRFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCD 468
QY 301 GMAADCTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDSVNDCGNSDEGSCSPAQTFRC 360
DB 469 GMAADCTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDSVNDCGNSDEGSCSPAQTFRC 528
QY 361 SNGKCLSKSQOQCKGDKDCGDSDEASCPRKVVVYCTKHYYRCINGLCLSKGNPECDGKED 420
DB 529 SNGKCLSKSQOQCKGDKDCGDSDEASCPRKVVVYCTKHYYRCINGLCLSKGNPECDGKED 588
QY 421 CSDGSDKDCDCCGLRSTFRQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480
DB 589 CSDGSDKDCDCCGLRSTFRQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMVLV 648
QY 481 SAACHYIDRGRFYSPDTQMTAFGLHDOSQSAAPGVQERLRKRIISHPEFNDFTDYDI 540
DB 649 SAACHYIDRGRFYSPDTQMTAFGLHDOSQSAAPGVQERLRKRIISHPEFNDFTDYDI 708
QY 541 ALLELEKPAEYSSMWRPCLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 600
DB 709 ALLELEKPAEYSSMWRPCLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 768
QY 601 QTTCEMLPQOITPRMNCVGFISGVNDCGDSGGLSSVEADGRITQAGVSWGDCQAQ 660
DB 769 QTTCEMLPQOITPRMNCVGFISGVNDCGDSGGLSSVEADGRITQAGVSWGDCQAQ 828
QY 661 RNKRGVYTRILPLFRDWIKENTGV 683
DB 828 RNKRGVYTRILPLFRDWIKENTGV 855
```

Db 829 RNKPGVYTRLPFRDWIKENTGV 851

RESULT 24  
US-10-296-115-1143  
; Sequence 1143, Application US/10296115  
; Publication No. US20040053248A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq Inc  
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides  
; FILE REFERENCE: 784PCT  
; CURRENT APPLICATION NUMBER: US/10/296.115  
; CURRENT FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: US09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; PRIOR APPLICATION NUMBER: US09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 1478  
; SEQ ID NO 1143  
; LENGTH: 851  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-296-115-1143

Query Match 98.8%; Score 3745; DB 15; Length 851;  
Best Local Similarity 99.0%; Pred. No. 3e-274;  
Matches 676; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRARSLSFVVTSVVAFPTDSKTVQRTQDSCSFGHLARGVELMRFTTPG 60  
Db 169 MAERVVMLPPRARSLSFVVTSVVAFPTDSKTVQRTQDSCSFGHLARGVELMRFTTPG 228  
Qy 61 PPDSPPYPAHARCCQALRGDADSVLSLFRSFDLASCDERSDLYTVYNTLSPMEPHALVQ 120  
Db 229 PPDSPPYPAHARCCQALRGDADSVLSLFRSFDLASCDERSDLYTVYNTLSPMEPHALVQ 288  
Qy 121 LCGTYPSPYNTLTHSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLRKAQGTFN 180  
Db 289 LCGTYPSPYNTLTHSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLRKAQGTFN 348  
Qy 181 SPYYPGHYPNIDCTWNIIEVNNQHVKKRFYLLERPRACGTCPCXDYVEINKEKCGE 240  
Db 349 SPYYPGHYPNIDCTWNIIEVNNQHVKKRFYLLERPRACGTCPCXDYVEINKEKCGE 408  
Qy 241 RSQGVVTSNSNKITVRHSDQSYTDGTGFLAELYSYDSSDPCPGQFTCRGCRIRKELRCD 300  
Db 409 RSQGVVTSNSNKITVRHSDQSYTDGTGFLAELYSYDSSDPCPGQFTCRGCRIRKELRCD 468  
Qy 301 GMADCTHSDDELNSCDAGHFTCKNFKCPLFWVCDSDVNDGNSDEOGSCPAOTPRC 360  
Db 469 GMADCTHSDDELNSCDAGHFTCKNFKCPLFWVCDSDVNDGNSDEOGSCPAOTPRC 528  
Qy 361 SNGKCLSKSQCGNKGDCGDSDEASCPKVVVVTCTKTYRCLNGLCLSKNPECDGKED 420  
Db 529 SNGKCLSKSQCGNKGDCGDSDEASCPKVVVVTCTKTYRCLNGLCLSKNPECDGKED 588  
Qy 421 CSDGSDKDCDCGRLSFTROARVVGCTDADGEMPMOVSLLHALCGHICGASLISPMVLV 480  
Db 589 CSDGSDKDCDCGRLSFTROARVVGCTDADGEMPMOVSLLHALCGHICGASLISPMVLV 648  
Qy 481 SAAACVYDDRGFRSDPTOWTAFIGLHPOSORSA PGUVERLKRILISHPPNDTTPVDI 540  
Db 649 SAAACVYDDRGFRSDPTOWTAFIGLHPOSORSA PGUVERLKRILISHPPNDTTPVDI 708  
Qy 541 ALLELEKPAEYSNVVRPCLPDAASHVFPAGKAIWVTGHTOYGATGALLIQKEIRVIN 600  
Db 709 ALLELEKPAEYSNVVRPCLPDAASHVFPAGKAIWVTGHTOYGATGALLIQKEIRVIN 768  
Qy 601 OTTENTLIPQOITPRMNCVGLSGVDSCGDSGGPLSSVEADGRIFQAGVSVNGDCAQ 660  
Db 769 OTTENTLIPQOITPRMNCVGLSGVDSCGDSGGPLSSVEADGRIFQAGVSVNGDCAQ 828  
Qy 661 RNKPGVYTRLPFRDWIKENTGV 683

Db 829 RNKPGVYTRLPFRDWIKENTGV 851

RESULT 25  
US-10-097-340-312  
; Sequence 312, Application US/10097340  
; Publication No. US20030087250A1  
; GENERAL INFORMATION:  
; APPLICANT: John MONAHAN  
; APPLICANT: Manjula GANNANARAPU  
; APPLICANT: Sebastian HOERSCH  
; APPLICANT: Shubhangi KAMATKAR  
; APPLICANT: Steve G. KOVATS  
; APPLICANT: Rachel E. MEYERS  
; APPLICANT: Michael MORRISSEY  
; APPLICANT: Peter OLANDT  
; APPLICANT: Ami SEN  
; APPLICANT: Peter VEIBY  
; APPLICANT: Gordon B. MILLS  
; APPLICANT: Robert C. BAST, Jr.  
; APPLICANT: Karen LU  
; APPLICANT: Rosemarie SCHMANDT  
; APPLICANT: Xumei ZHAO  
; APPLICANT: Karen GLATT  
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer  
; FILE REFERENCE: MRI-030  
; CURRENT APPLICATION NUMBER: US/10/097,340  
; CURRENT FILING DATE: 2002-03-14  
; PRIOR APPLICATION NUMBER: 60/276,025  
; PRIOR FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 60/325,149  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: 60/276,026  
; PRIOR FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 60/324,967  
; PRIOR FILING DATE: 2001/09/26  
; PRIOR APPLICATION NUMBER: 60/311,732  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/325,102  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: 60/323,580  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 363  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 312  
; LENGTH: 782  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-097-340-312

Query Match 86.0%; Score 3259.5; DB 14; Length 782;  
Best Local Similarity 88.4%; Pred. No. 1.3e-237;  
Matches 604; Conservative 1; Mismatches 5; Indels 73; Gaps 2;

Qy 1 MAERVVMLPPRARSLSFVVTSVVAFPTDSKTVQRTQDSCSFGHLARGVELMRFTTPG 60  
Db 173 MAERVVMLPPRARSLSFVVTSVVAFPTDSKTVQRTQDSCSFGHLARGVELMRFTTPG 232  
Qy 61 PPDSPPYPAHARCCQALRGDADSVLSLFRSFDLASCDERSDLYTVYNTLSPMEPHALVQ 120  
Db 233 PPDSPPYPAHARCCQALRGDADSVLSLFRSFDLASCDERSDLYTVYNTLSPMEPHALVQ 292  
Qy 121 LCGTYPSPYNTLTHSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLRKAQGTFN 180  
Db 293 LCGTYPSPYNTLTHSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLRKAQGTFN 352  
Qy 181 SPYYPGHYPNIDCTWNIIEVNNQHVKKRFYLLERPRACGTCPCXDYVEINKEKCGE 240  
Db 353 SPYYPGHYPNIDCTWNIIEVNNQHVKKRFYLLERPRACGTCPCXDYVEINKEKCGE 412  
Qy 241 RSQGVVTSNSNKITVRHSDQSYTDGTGFLAELYSYDSSDPCPGQFTCRGCRIRKELRCD 300

```
Db 413 RSQFVVTNSNKTITVRHSDQSYDTGFLAEYLSYDSSD----- 451
Qy 301 GMAIDCTHSDBELNCSGAGHOFTCKNKECPLEFVWCDSVNDGDNDEGSCSPAOTFRC 360
Db 452 -----RCAGHOFCKNFKCKPLFVWCDSVNDGDNSEOGC----- 488
Qy 361 SNGKCLSKSQOQNGKXDCGSDGSDBASCPKVVVVTCTKHTYRCUNGCLCLSKGNPECDKED 420
Db 489 -----NMVVVTCTGHTYRCUNGCLCLSKGNPECDKED 519
Qy 421 CSQGSDEKDCDCGLRSFTQARVVGCTDADBEGEMPMQVSLHALGQGHICGASLISPMWLV 480
Db 520 CSQGSDEKDCDCGLRSFTQARVVGCTDADBEGEMPMQVSLHALGQGHICGASLISPMWLV 579
Qy 481 SAACHYIDDRGFPRYSPTQMTAFGLHDOSRASPQVQERLKRITISHPFNDFTFDYDI 540
Db 580 SAACHYIDDRGFPRYSPTQMTAFGLHDOSRASPQVQERLKRITISHPFNDFTFDYDI 639
Qy 541 ALLELEKPAEYSSWVRPICI.PDASHVFPAGKAIWVTGMHTQYGGTGALLILQKEIRVIN 600
Db 640 ALLELEKPAEYSSWVRPICI.PDASHVFPAGKAIWVTGMHTQYGGTGALLILQKEIRVIN 699
Qy 601 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSGGPLSYBADGRITQAGVSWGDCQAQ 660
Db 700 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSGGPLSYBADGRITQAGVSWGDCQAQ 759
Qy 661 RNKRGVYTRLP.LFRDWMIKENTGV 683
Db 760 RNKRGVYTRLP.LFRDWMIKENTGV 782
```

## RESULT 26

```
US-09-900-751-2
; Sequence 2, Application US/09900751
; Patent No. US2002026653A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING SERINE
; TITLE OF INVENTION: PROTEASE GENE DISRUPTIONS
; FILE REFERENCE: R-386
; CURRENT APPLICATION NUMBER: US/09/900, 751
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/217,449
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/223,170
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 60/223,460
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-900-751-2
```

Query Match 84.7%; Score 3209; DB 9; Length 855;

Best Local Similarity 83.2%; Pred. No. 9.7e-234; Matches 568; Conservative 59; Mismatches 56; Indels 0; Gaps 0;

```
Qy 1 MAERVVMLPPRARSLKSFVTVSVAPPTDSKTQVQRQDNSSCGFGLHARGVELMARFTTPG 60
Db 173 MAERVVMLPPRARSLKSFVTVSVAPPTDSKTQVQRQDNSSCGFGLHARGVELMARFTTPG 232
Qy 61 FPDSPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPEMHALVQ 120
Db 233 FPDSPYPAHARCOMALRGDADSVLSTFRSFDVAPCEHGSIDLVTYNTLSPEMHALVVR 292
Qy 121 LCGTSPSYNLTHSSGNVLLITLITNTERRHRRFETAFQOLPRMSSCGRLKKAQCTEN 180
Db 293 LCGTSPSYNLTHSSGNVLLITLITNTERRHRRFETAFQOLPRMSSCGRLKKAQCTEN 352
```

```
Qy 181 SPYPCGHYPENIDCTNNIEVPNNQHVKRFKFFYLLEPRACGTCRKYVEINGEKYCGE 240
Db 353 SPYPCGHYPENIDCTNNIEVPNNQHVKRFKFFYLLEPRACGTCRKYVEINGEKYCGE 412
Qy 241 RSQFVVTNSNKTITVRHSDQSYDTGFLAEYLSYDSSDPCPQOFTCRGRCTRKLRCQ 300
Db 413 RSQFVVTNSNKTITVRHSDQSYDTGFLAEYLSYDSSDPCPQOFTCRGRCTRKLRCQ 472
Qy 301 GMAIDCTHSDBELNCSGAGHOFTCKNKECPLEFVWCDSVNDGDNDEGSCSPAOTFRC 360
Db 473 GMAIDCTHSDBELNCSGAGHOFTCKNKECPLEFVWCDSVNDGDNDEGSCSPAOTFRC 532
Qy 361 SNGKCLSKSQOQNGKXDCGSDGSDBASCPKVVVVTCTKHTYRCUNGCLCLSKGNPECDKED 420
Db 533 SNGKCLPQSGKNGKXDCGSDGSDBASCPKVVVVTCTKHTYRCUNGCLCLSKGNPECDKED 592
Qy 421 CSQGSDEKDCDCGLRSFTQARVVGCTDADBEGEMPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSQGSDEKDCDCGLRSFTQARVVGCTDADBEGEMPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAACHYIDDRGFPRYSPTQMTAFGLHDOSRASPQVQERLKRITISHPFNDFTFDYDI 540
Db 653 SAACHYIDDRGFPRYSPTQMTAFGLHDOSRASPQVQERLKRITISHPFNDFTFDYDI 712
Qy 541 ALLELEKPAEYSSWVRPICI.PDASHVFPAGKAIWVTGMHTQYGGTGALLILQKEIRVIN 600
Db 713 ALLELEKPAEYSSWVRPICI.PDASHVFPAGKAIWVTGMHTQYGGTGALLILQKEIRVIN 772
Qy 601 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSGGPLSYBADGRITQAGVSWGDCQAQ 660
Db 773 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSGGPLSYBADGRITQAGVSWGDCQAQ 832
Qy 661 RNKRGVYTRLP.LFRDWMIKENTGV 683
Db 833 RNKRGVYTRLP.LFRDWMIKENTGV 855
```

## RESULT 27

```
US-10-072-012-355
; Sequence 355, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchervet, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patnursajan, Meera
; APPLICANT: Shimkete, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Baha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
```

```

PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See file Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 355
LENGTH: 855
TYPE: PRT
ORGANISM: Mus musculus
US-10-072-012-355

Query Match      84.7%  Score 3209; DB 15; Length 855;
Best Local Similarity 83.2%  Pred. No. 9.7e-234;
Matches 568; Conservative 59; Mismatches 56; Indels 0; Gaps 0;

QY 1 MAERVYMLPPRARSLKSPVTVSVAPPTDSKTVOPTODNSCSFGLHARGVELMRFTTPG 60
DB 173 MAERVYMLPPRARSLKSPVTVSVAPPTDSKTVOPTODNSCSFGLHARGVELMRFTTPG 232
QY 61 FPDSPYAHARQWALRGDADSVLSTFRSFDLASCDERGSIDLTVYNTLSPMEPHALVQ 120
DB 233 FPDSPYAHARQWALRGDADSVLSTFRSFDVAPCEHSGDLTVYNTLSPMEPHALVQ 292
QY 121 LCGTYPSSYNLTFHSSQNVLLITLITERRHGPPEATFQPLPMSSCGGLRAQGTEN 180
DB 293 LCGTYPSSYNLTFHSSQNVLLITLITERRHGPPEATFQPLPMSSCGGLRAQGTEN 352
QY 181 SPVYHGPVPIIDCTNMIIEVNNHNVYRFEFLYLBPRACGCPDYVEINEKXCGE 240
DB 353 SPVYHGPVPIIDCTNMIIEVNNHNVYRFEFLYLBPRACGCPDYVEINEKXCGE 412
QY 241 RSQFVTVSNKLTIVRFHSDQSYDTGTFLAEYLSYDSSDPCGQFTGRTGCIKEIRCD 300
DB 413 RSQFVTVSNKLTIVRFHSDQSYDTGTFLAEYLSYDSSDPCGQFTGRTGCIKEIRCD 472
QY 301 GMADCTHSDDELNCSGAGHOFCTCKNFKCKPLFVNCDSVNDGNSDBQGCCCPAQTPRC 360
DB 473 GMADCTHSDDELNCSGAGHOFCTCKNFKCKPLFVNCDSVNDGNSDBQGCCCPAQTPRC 532
QY 473 GMADCTHSDDELNCSGAGHOFCTCKNFKCKPLFVNCDSVNDGNSDBQGCCCPAQTPRC 532
DB 361 SNGKCLSKSQOQCKNDKCGGSDASCPKXNVVCTHTATYCLNGLCLSKNPECDKED 420
QY 533 SNGKCLSKSQOQCKNDKCGGSDASCPKXNVVCTHTATYCLNGLCLSKNPECDKED 592
DB 421 CSDSDEKDCDCGARSFTQARVVGCTDADGEMPMQVSLHALQGHICGSLSPMWLV 480
QY 593 CSDSDEKDCDCGARSFTQARVVGCTDADGEMPMQVSLHALQGHICGSLSPMWLV 652
DB 481 SAHACFYIDDRGFRYSPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPFNDFTFYDI 540
QY 653 SAHACFYIDDRGFRYSPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPFNDFTFYDI 712
DB 541 ALLELEKPAEYSSNVPRICLPDASHVPAGKAIWVTGNGHYOGTGLILIOKEIRVIN 600
QY 713 ALLELEKPAEYSSNVPRICLPDASHVPAGKAIWVTGNGHYOGTGLILIOKEIRVIN 772
DB 601 QTTCENLIPQOITPRMNCVGFISGVDSCQDSCGSLSSVADGRIFQAGVSVNGDCAQ 660
QY 773 QTTCENLIPQOITPRMNCVGFISGVDSCQDSCGSLSSVADGRIFQAGVSVNGDCAQ 832
DB 660 QTTCENLIPQOITPRMNCVGFISGVDSCQDSCGSLSSVADGRIFQAGVSVNGDCAQ 832
```

```

QY 661 RNKGYVTRLPFRDWMKENTGV 683
DB 833 RNKGYVTRLPFRDWMKENTGV 855

RESULT 28
US-10-072-012-413
Sequence 413, Application US/10072012
Publication No. US2004003493A1
GENERAL INFORMATION:
APPLICANT: Tchernov, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zethusen, Bryan
APPLICANT: Paturajan, Meera
APPLICANT: Shukets, Richard
APPLICANT: Li, Li
APPLICANT: Gangolli, Baha
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Coleman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OR INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See file Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 413
LENGTH: 855
TYPE: PRT
ORGANISM: Mus musculus
US-10-072-012-413

Query Match      84.7%  Score 3209; DB 15; Length 855;
Best Local Similarity 83.2%  Pred. No. 9.7e-234;
Matches 568; Conservative 59; Mismatches 56; Indels 0; Gaps 0;

QY 1 MAERVYMLPPRARSLKSPVTVSVAPPTDSKTVOPTODNSCSFGLHARGVELMRFTTPG 60
DB 173 MAERVYMLPPRARSLKSPVTVSVAPPTDSKTVOPTODNSCSFGLHARGVELMRFTTPG 232
```

Dd	173	NAVERVTLTPRABRLAKSVLTVSVAFPLDPRMLQOTQDNQSCGFALHAHAATRRFTTPG	232
Oy	61	PPDSFYPAHARQOMALRGDADSVLSTFSPFDIASCDERSGDLVTTYNTLSPEPHALVQ	120
Dd	233	PENSFYPAHARQOMVLRGDADSVLSTLFPSFDVAPCDHEHSDLVTVYDLSLSPMBPHAVR	232
Oy	121	LCGTVPSPNLTFFHSSQNVLLTLLINTBRHRHGFEPATFPQLPRMSCGRLKKAQOTFN	180
Dd	233	LCGTSPSPNLTFFLSQNVFLVTLINTDRHNGFEATFPQLPRMSCGGFLDQTGTFPS	352
Oy	181	SPYFPGHYPNIDCTWNIEVPNNQHVAPKPFYLLPERRACSTCPDYVEINGEKYCGE	240
Dd	353	SPYFEGHYPNINCTWNIKVPPNNRNVKVPFKLYLVDPNVPVSGCTKDYEINGEKYCGE	412
Oy	241	RSQFVLTSSNMTIYRHFHSDOSTDTPDGLAEYLSYSSSDPCQPOFCRTRRCIRKELRCD	300
Dd	413	RSQFVSSSSSKITVHFHSDHSTDTGFLAEYLSYSSNDPCPMFMCKTGRCTRKELRCD	472
Oy	301	GWADCTDHSDELNCSCDAGHOFTCKRNKFCPLFWVCDSVNDCGDNSDQSCSPAQTFRC	360
Dd	473	GWADCPDYDSERYCRGNATHQFTCKNQFCKPLFWVCDSVNDCGDSDBEGSCSPAGSFKC	522
Oy	361	SNGKCLTSKQOQNGKDCDGDGDSDEASCPRXVNTYCTKHYRCLNGLCLSGANPECDKED	420
Dd	533	SNGKCLPOSQKQNGKDNCGDSDSEASCDSNVVYSCTKYTRCQNGCLCSGANECDGKTD	592
Oy	421	CSDGDEKDCDCGLRSFTRQARVVGSTDADEGEMPMQVSLHAUGOGHICGASLSIPMWLV	480
Dd	593	CSDGDEKDCDCGLRSFTTQARVVGSTNADGEMPMQVSLHAUGOGHICASLSIPMWLV	652
Oy	481	SAHACYTIDDRGRYSYDPTQMTAFLGLHDSQRSAPVQOEHRRLKRIISHPEPNDFTEDYDI	540
Dd	653	SAAHFOQDDKNKFSYDTWMTAFLGLHDSKKSASGVQELKRIITHPSFNDFTFEDYDI	712
Oy	541	ALLETLEKPAEVSMMVAPICLPASHYFPPAGKAIWMTGMGTQVGGTGAALLQGEIRVIN	600
Dd	713	ALLETLEKSVYSITVAPICLPDATHVFPAGKAIWMTGMGTKEGGTGALLQGEIRVIN	772
Oy	601	QTTCENTLFPQOITPRMMCVGLSGVDSCQDSDGCGPLSSVEADGRITFOAGVSMGDCAQ	660
Dd	773	QTTCEDLMPQOITPRMMCVGLSGVDSCQDSDGCGPLSSLEKXGRMFQAGVSMGDCAQ	832
Oy	661	RNRKGVYTRLPLFRDMIKENTGV	683
Dd	833	RNRKGVYTRLPVVRWDIKENTGV	855
RESULT 29			
US-10-072-012-356			
Sequence 356, Application US/10072012			
Publication No. US20040033493A1			
GENERAL INFORMATION:			
APPLICANT: Tchernev, Velizar			
APPLICANT: Spytek, Kimberly			
APPLICANT: Zehnusen, Bryan			
APPLICANT: Patturejan, Meera			
APPLICANT: Shimets, Richard			
APPLICANT: Li, Li			
APPLICANT: Gangolli, Bsha			
APPLICANT: Padigaru, Muralidhara			
APPLICANT: Anderson, David W.			
APPLICANT: Rastelli, Luca			
APPLICANT: Miller, Charles E.			
APPLICANT: Gerlach, Valerie			
APPLICANT: Taupier Jr, Raymond J.			
APPLICANT: Gusev, Vladimir Y.			
APPLICANT: Colman, Steven D.			
APPLICANT: Wolenc, Adam R.			
APPLICANT: Pena, Carol E. A			
APPLICANT: Futrak, Katarzyna			
APPLICANT: Grosse, William M.			
APPLICANT: Alsdbrook II, John P.			
APPLICANT: Lepley, Denise M.			

	APPLICANT: Rieger, Daniel K.	
	TITLE OF INVENTION: Burgess, Catherine E.	
	FILE REFERENCE: 21402-258	
	CURRENT APPLICATION NUMBER: US/10/072.012	
	CURRENT FILING DATE: 2002-01-31	
	PRIOR APPLICATION NUMBER: 60/265,102	
	PRIOR FILING DATE: 2001-01-30	
	PRIOR APPLICATION NUMBER: 60/265,514	
	PRIOR FILING DATE: 2001-01-31	
	PRIOR APPLICATION NUMBER: 60/265,517	
	PRIOR FILING DATE: 2001-01-31	
	PRIOR APPLICATION NUMBER: 60/265,412	
	PRIOR FILING DATE: 2001-01-31	
	PRIOR APPLICATION NUMBER: 60/265,395	
	PRIOR FILING DATE: 2001-01-31	
	PRIOR APPLICATION NUMBER: 60/266,406	
	PRIOR FILING DATE: 2001-02-02	
	PRIOR APPLICATION NUMBER: 60/266,767	
	PRIOR FILING DATE: 2001-02-05	
	PRIOR APPLICATION NUMBER: 60/267,057	
	PRIOR FILING DATE: 2001-02-07	
	PRIOR APPLICATION NUMBER: 60/266,975	
	PRIOR FILING DATE: 2001-02-07	
	PRIOR APPLICATION NUMBER: 60/267,459	
	PRIOR FILING DATE: 2001-02-08	
	Remaining Prior Application data removed - See File Wrapper or PALM.	
	NUMBER OF SEQ ID NOS: 1391	
	SOFTWARE: PatentIn Ver. 2.1	
	SEQ ID NO 356	
	LENGTH: 855	
	TYPE: PRT	
	ORGANISM: Rattus norvegicus	
	US-10-072-012-356	
QY	Query Match	84.4%; Score 3197; DB 15; Length 855;
Db	Best Local Similarity	83.0%; Pred.No.7.8e-233;
	Matches	567; Conservative 59; Mismatches 57; Indels 0; Gaps 0
QY	1 MAERVWMLPRASLSKFSVVTSVAFPTDSKVORTDNSCSFELHARGVELMRFTTGG	60
Db	173 MAVERVVTLPPRARALMSFLVLTSAVFIDRPMQRITDNGSFLAHRRGRVTRFTTGG	232
QY	61 FPDSFYPAHARCOWMLRGDADSVSLSTRSTDLASCDESGDVLTVYNTLSPMEHALVQ	120
Db	233 FPNSEYPHAACOWWLREGDADVSLSTRSPDVAPCDGHDLDLVYVDSLSEMEHAVR	292
QY	121 LCGTPPSPNLTFFSSQWVLLITLIITNERHHGPGEATFPOLPRMSSCGRLRKQGTEN	180
Db	293 LCGTFPSBYNLTFSSQWVLTVLITINDRHHPGEATFPOLPRMSSCGGLSEAGTFS	352
QY	181 SPYYPGHYPNIDCTWNIEVNNOHVKKRFXYLLEBRRAAGTCPKDYVEINSEKYCE	240
Db	353 SPYYPGHYPPINICTWNIKVPNNRNVRKFKLFYLVDENIPVGSCTKDYVEINGEKFCGE	412
QY	241 RSOPVYNSNKITRYREHSDDSYTDGTGLAYLVSDSDPCGCQTCTRCIRKELRCD	300
Db	413 RSQFVSSNSSKITVHFSDHSYDTDTGFLEAELYLSDNDPCGMFMCKTGRIKDKLCD	472
QY	301 GMADCTHSDSLNCSCDAGHQFTCKNKPCKPLFWVCDSVNCGDNSDEOGSCSPAOTERC	360
Db	473 GMADCPDVSDBRHRCRNATHQFMCNQCXKPLFWVCDSVNCGDNSDEOGSCSPAOSFC	532
QY	361 SNGKLKSQQCNKGDCDGSDSEASCPKXNVVTTCTHTYCLNGLCLSKNPEDCGKD	420
Db	533 SNGKLCPQSQQCNKGDCDGSDSEASCNVNAVASCTKYTYRCQNGLCINKNPEDCGKXD	592
QY	421 CSDSDKDDCCGRLSFTRORARYGGTPADBGEMPOVSLHALGGSHLGASLISPMULV	480
Db	593 CSDSDERAKCUGLRSFTKQRPVGGTNADGEMPOVSLHALGGSHLGASLISPMLV	652
QY	481 SAACGYIDNRGRYSPTQWTAFILGHDOGSARAGVGERRLKRIISHPFNDFETDYDI	540

Db 653 SAACFODETIFKYS DHTMTAFGLDQSKRSASGVQEHKLRITHPSFNDFTPDYDI 712  
Qy 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGHTGHTQYGGTALLIQKEIRVIN 600  
Db 713 ALLELEKPAEYSTVRPCLPDNTHVPAGKAIWVTGHTGHTGCTGALLIQKEIRVIN 772  
Qy 601 OTTCENLLPQOITPRMNCVGLSGVDS CGDSGGLSSVADGRIFQAGVWSMGDCAQ 660  
Db 773 OTTCENLLPQOITPRMNCVGLSGVDS CGDSGGLSSVADGRIFQAGVWSMGDCAQ 832  
Qy 661 RNKPGVYTRLP LFRDWMKENTGV 683  
Db 833 RNKPGVYTRIP EYRDMWKEQTGV 855

RESULT 30  
US-10-072-012-414  
Sequence 414, Application US/10072012  
Publication No. US20040033493A1  
GENERAL INFORMATION:  
APPLICANT: Tchenev, Velizar  
APPLICANT: Spytek, Kimberly  
APPLICANT: Zernhusen, Bryan  
APPLICANT: Paturajan, Meera  
APPLICANT: Shimkets, Richard  
APPLICANT: Li, Li  
APPLICANT: Gangolli, Esha  
APPLICANT: Padigaru, Murallidhara  
APPLICANT: Anderson, David W.  
APPLICANT: Rastelli, Luca  
APPLICANT: Miller, Charles E.  
APPLICANT: Taupier Jr, Raymond J.  
APPLICANT: Gusev, Vladimir Y.  
APPLICANT: Coleman, Steven D.  
APPLICANT: Wolenc, Adam R.  
APPLICANT: Pena, Carol E. A  
APPLICANT: Furtak, Katarzyna  
APPLICANT: Grose, William M.  
APPLICANT: Alsbrook II, John P.  
APPLICANT: Lepley, Denise M.  
APPLICANT: Rieger, Daniel K.  
APPLICANT: Burgess, Catherine E.  
TITLE OR INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-258  
CURRENT APPLICATION NUMBER: US/10/072.012  
PRIOR FILING DATE: 2002-01-31  
PRIOR APPLICATION NUMBER: 60/265,102  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: 60/265,514  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,517  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,412  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/265,395  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: 60/266,406  
PRIOR FILING DATE: 2001-02-02  
PRIOR APPLICATION NUMBER: 60/266,767  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: 60/267,057  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/266,975  
PRIOR FILING DATE: 2001-02-07  
PRIOR APPLICATION NUMBER: 60/267,459  
PRIOR FILING DATE: 2001-02-08  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1391  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 414  
LENGTH: 855  
TYPE: PRT

ORGANISM: Rattus norvegicus  
US-10-072-012-414  
Query Match 84.4%; Score 3197; DB 15; Length 855;  
Best Local Similarity 83.0%; Pred. No. 7.8e-23;  
Matches 567; Conservative 59; Mismatches 57; Indels 0; Gaps 0;  
Qy 1 MAERVYMLPPRARSLKPFVTSVAPPTDQRTQODNSCFGLHARGVELARFTTPG 60  
Db 173 MAERVYMLPPRARSLKPFVTSVAPPTDQRTQODNSCFGLHARGVELARFTTPG 232  
Qy 61 FPDSPYPAHARCCQWALRGDADSVLSLTFRSFDLASCDERSDLYVTYNTLSPEPHALVQ 120  
Db 233 FPDSPYPAHARCCQWALRGDADSVLSLTFRSFDVAPCDGHDSDLYVTYNTLSPEPHALVYR 292  
Qy 121 LCGTYPPSYNLTFHSSQNVLLITLINTERRHGFETFFQLPRMSSCGRLKAQGTEN 180  
Db 293 LCGTFSPSYNLTFHSSQNVFLVTLITNTDRHGFETFFQLPRMSSCGRLSAQGTFS 352  
Qy 181 SPVYRGHYPPNIDCTNMI EVPNNOHVVRPKFYLEPRRACGTCPDYVEINGEYKCGE 240  
Db 353 SPVYRGHYPPNIDCTNMIKVNNNVKVRKFLFLVDPNIPVSGCTQDYVEINGEYKCGE 412  
Qy 241 RSQFVYTSNKNKLTFRSHSDSYTDGTFLA EYLSYDSDPCPGQFTCRTRCIRKELRCD 300  
Db 413 RSQFVYTSNKNKLTFRSHSDSYTDGTFLA EYLSYDSDPCPGFMCKTRCIRKDLRCD 472  
Qy 301 GMADCTHSDBLNCSCAHQFTCKNFKCPLTFVNCDSVNDCCGNSDEGSCCPAQTFRC 360  
Db 473 GMADCPDYSBRHRCNATHQFMCKNCFCKPLFVNCDSVNDCCGNSDEGSCCPAGSFKC 532  
Qy 361 SNGKCLSKSOCCNCKDCCGSDGSDASCPKNVYCTGHTYRCLNGCLSKGNPBCDCKED 420  
Db 533 SNGKCLSKSOCCNCKDCCGSDGSDASCDVNAVASCTYTYRCCNGCLSKNGNPECDCKED 592  
Qy 421 CSDGSDKDCDCGLRSTFRQARVVGCTDADEGEPMQVSLHALGQGHICGASLSPNMLV 480  
Db 593 CSDGSDKDCDCGLRSTFRQARVVGCTDADEGEPMQVSLHALGQGHICGASLSPNMLV 652  
Qy 481 SAACIYIDDRGFRYSPTQNTAFGLHDOSORSAPGVQERLKRITHPSFNDFTPDYDI 540  
Db 653 SAACFODETIFKYS DHTMTAFGLDQSKRSASGVQEHKLRITHPSFNDFTPDYDI 712  
Qy 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGHTGHTGCTGALLIQKEIRVIN 600  
Db 713 ALLELEKPAEYSTVRPCLPDNTHVPAGKAIWVTGHTGHTGCTGALLIQKEIRVIN 772  
Qy 601 OTTCENLLPQOITPRMNCVGLSGVDS CGDSGGLSSVADGRIFQAGVWSMGDCAQ 660  
Db 773 OTTCENLLPQOITPRMNCVGLSGVDS CGDSGGLSSVADGRIFQAGVWSMGDCAQ 832  
Qy 661 RNKPGVYTRLP LFRDWMKENTGV 683  
Db 833 RNKPGVYTRIP EYRDMWKEQTGV 855

Search completed: November 29, 2004, 08:49:50  
Job time : 94.8166 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: November 29, 2004, 08:20:59 ; Search time 25.7568 Seconds  
(without alignments)  
1758.570 Million cell updates/sec

Title: US-09-936-333-5  
Perfect score: 3789  
Sequence: 1 MAERVYMLPPRARSLSKFV.....PGVYTRLPFRDMIKENTGV 683

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3754	99.1	855	2	US-09-027-337-2
2	3754	99.1	855	4	US-09-644-600-2
3	3754	99.1	855	4	US-09-654-600A-2
4	3123	82.4	902	4	US-09-644-600A-10
5	3123	82.4	902	4	US-09-654-600A-10
6	1319	34.8	241	4	US-09-657-968B-2
7	703.5	18.6	798	1	US-08-200-900A-2
8	703.5	18.6	798	4	US-08-794-042-2
9	703.5	18.6	798	5	PCT-US94-00616-2
10	588	15.5	407	3	US-09-734-675-4
11	560.5	14.8	492	4	US-09-685-166A-895
12	560.5	14.8	492	4	US-09-879-792-14
13	558.5	14.7	492	4	US-09-679-426-895
14	558.5	14.7	492	3	US-09-342-749-2
15	558.5	14.7	492	4	US-09-691-840-2
16	547.5	14.4	235	3	US-08-944-483-65
17	544.5	14.4	235	3	US-08-807-151-3
18	544.5	14.4	235	3	US-09-478-957-3
19	543.5	14.3	454	3	US-09-518-046-2
20	527	13.9	466	4	US-10-177-661-4
21	526.5	13.9	638	2	US-08-681-151-3
22	521.5	13.8	418	4	US-10-177-661-6
23	518.5	13.7	477	4	US-10-177-661-2
24	518.5	13.7	562	4	US-09-879-792-12
25	512.5	13.5	283	3	US-08-807-151-1
26	512.5	13.5	283	3	US-09-478-957-1
27	511.5	13.5	418	1	US-08-508-448C-25

28	511.5	13.5	418	4	US-09-370-838-82	Sequence 82, Appl
29	511.5	13.5	418	4	US-09-370-838-83	Sequence 83, Appl
30	511.5	13.5	418	4	US-09-854-133-82	Sequence 82, Appl
31	511.5	13.5	418	4	US-09-854-133-83	Sequence 83, Appl
32	510.5	13.5	418	4	US-09-370-838-62	Sequence 62, Appl
33	510.5	13.5	418	4	US-09-854-133-62	Sequence 62, Appl
34	505.5	13.3	376	4	US-09-820-002-2	Sequence 2, Appl
35	504	13.3	790	4	US-08-991-761A-13	Sequence 13, Appl
36	500.5	13.2	546	6	5200340-6	Patent No. 5200340
37	497.5	13.1	222	1	US-08-508-448C-19	Sequence 19, Appl
38	496	13.1	416	2	US-09-000-846-2	Sequence 2, Appl
39	493.5	13.0	791	1	US-08-643-219-1	Sequence 1, Appl
40	493.5	13.0	791	3	US-08-851-350-1	Sequence 1, Appl
41	492.5	13.0	583	4	US-09-976-594-837	Sequence 837, App
42	492.5	13.0	790	1	US-08-469-486-54	Sequence 54, Appl
43	492.5	13.0	790	2	US-08-469-658-54	Sequence 54, Appl
44	492	13.0	417	4	US-09-820-002-4	Sequence 4, Appl
45	491.5	13.0	791	2	US-09-131-995-1	Sequence 1, Appl
46	491.5	13.0	791	3	US-08-832-087B-1	Sequence 1, Appl
47	491.5	13.0	791	3	US-09-132-154-1	Sequence 1, Appl
48	491.5	13.0	791	4	US-08-991-761A-6	Sequence 6, Appl
49	491.5	13.0	791	4	US-08-924-287A-1	Sequence 1, Appl
50	491.5	13.0	810	1	US-07-854-603-2	Sequence 2, Appl
51	491.5	13.0	810	1	US-08-147-000B-29	Sequence 29, Appl
52	491.5	13.0	810	3	US-09-086-514-1	Sequence 1, Appl
53	491.5	13.0	810	4	US-09-192-012-5	Sequence 5, Appl
54	491.5	13.0	810	4	US-09-403-736-1	Sequence 1, Appl
55	491.5	13.0	810	4	US-09-701-265-1	Sequence 1, Appl
56	491.5	13.0	810	6	5200340-8	Patent No. 5200340
57	491.5	13.0	812	1	US-08-248-629A-1	Sequence 1, Appl
58	491.5	13.0	812	1	US-08-451-932-1	Sequence 1, Appl
59	491.5	13.0	812	1	US-08-452-260-1	Sequence 1, Appl
60	491.5	13.0	812	1	US-08-326-785-1	Sequence 1, Appl
61	491.5	13.0	812	2	US-08-612-788-1	Sequence 1, Appl
62	491.5	13.0	812	2	US-08-605-598B-1	Sequence 1, Appl
63	491.5	13.0	812	2	US-08-429-743-1	Sequence 1, Appl
64	491.5	13.0	812	2	US-08-866-735-1	Sequence 1, Appl
65	491.5	13.0	812	3	US-09-066-028-1	Sequence 3, Appl
66	491.5	13.0	812	4	US-09-192-012-3	Sequence 1, Appl
67	491.5	13.0	812	4	US-09-335-325-1	Sequence 1, Appl
68	491.5	13.0	812	4	US-08-991-761A-12	Sequence 12, Appl
69	491.5	13.0	812	5	PCT-US95-05107-1	Sequence 1, Appl
70	488.5	12.9	814	1	US-08-750-711-1	Sequence 1, Appl
71	484.5	12.8	415	1	US-08-073-531B-1	Sequence 1, Appl
72	484.5	12.8	415	2	US-08-766-288-1	Sequence 1, Appl
73	484.5	12.8	461	3	US-08-742-877-2	Sequence 2, Appl
74	481.5	12.7	315	4	US-09-386-653A-9	Sequence 9, Appl
75	481.5	12.7	812	4	US-08-991-761A-7	Sequence 7, Appl
76	481.5	12.7	405	3	US-09-734-675-2	Sequence 2, Appl
77	481	12.7	810	4	US-08-991-761A-11	Sequence 11, Appl
78	480.5	12.7	461	3	US-09-053-871A-21	Sequence 21, Appl
79	480.5	12.7	461	4	US-10-133-907-5	Sequence 5, Appl
80	480.5	12.7	461	6	5521070-2	Patent No. 5521070
81	479.5	12.7	415	1	US-08-295-411-2	Sequence 2, Appl
82	479.5	12.7	415	5	US-08-955-471-2	Sequence 2, Appl
83	479.5	12.7	266	2	PCT-US97-10242-2	Sequence 2, Appl
84	478.5	12.6	266	4	US-09-644-600A-3	Sequence 3, Appl
85	478.5	12.6	266	4	US-09-654-600A-3	Sequence 3, Appl
86	478.5	12.6	266	4	US-09-027-337-3	Sequence 3, Appl
87	477	12.6	527	6	5520913-1	Patent No. 5520913
88	476.5	12.6	290	4	US-09-386-653A-7	Sequence 7, Appl
89	476.5	12.6	320	4	US-10-140-002-222	Sequence 222, App
90	476	12.6	328	4	US-09-386-642-11	Sequence 11, Appl
91	474.5	12.5	809	4	US-08-991-761A-9	Sequence 9, Appl
92	473.5	12.5	255	3	US-08-944-483-67	Sequence 67, Appl
93	473.5	12.5	415	4	US-09-118-748-2	Sequence 2, Appl
94	472	12.5	455	3	US-09-261-416-2	Sequence 2, Appl
95	472	12.5	527	2	US-08-811-949-39	Sequence 39, Appl
96	470	12.4	319	4	US-09-386-642-12	Sequence 12, Appl
97	469.5	12.4	352	6	5244676-5	Patent No. 5244676
98	469	12.4	472	2	US-08-811-949-63	Sequence 63, Appl
99	467.5	12.3	437	2	US-08-811-949-51	Sequence 51, Appl
100	467	12.3	527	1	US-07-609-510B-16	Sequence 16, Appl

## ALIGNMENTS

## RESULT 1

US-09-027-337-2  
; Sequence 2, Application US/09027337B  
; Patent No. 5972616  
; GENERAL INFORMATION:  
; APPLICANT: Tanimoto, Timothy J.  
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in  
; FILE REFERENCE: D6064  
; CURRENT APPLICATION NUMBER: US/09/027,337B  
; NUMBER OF SEQ ID NOS: 13  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides  
; Patent No. 5972616  
US-09-027-337-2

Query Match 99.1%; Score 3754; DB 2; Length 855;  
Best Local Similarity 99.3%; Pred. No. 3.9e-262;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNSCSFGIHAAGVEIMRFTTPG 60  
DB 173 MAERVVMLEPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNSCSFGIHAAGVEIMRFTTPG 232  
QY 61 FPDSPYPAHARCOMALRGADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120  
DB 233 FPDSPYPAHARCOMALRGADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292  
QY 121 LCGTTPPSYNLTFHSSQNVLLITLINTERRRHGFEEATFFOLPRMSSCGRLRKAQGTEN 180  
DB 293 LCGTTPPSYNLTFHSSQNVLLITLINTERRRHGFEEATFFOLPRMSSCGRLRKAQGTEN 352  
QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVAVRPFKFFYLLEPRRACGTCPKDYVEINGEKYCGE 240  
DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVAVRPFKFFYLLEPRRACGTCPKDYVEINGEKYCGE 412  
QY 241 RSQFVVTSSNKTITVRFHSDQSYTDTGFLAELYSYSSDPCPGQFTCRGRICRKELRCD 300  
DB 413 RSQFVVTSSNKTITVRFHSDQSYTDTGFLAELYSYSSDPCPGQFTCRGRICRKELRCD 472  
QY 301 GMADCTDHSDELNCSCDAGHOFCKNPKFCKPLFWVCDVNDGDNSEGGSCPAQTFRC 360  
DB 473 GMADCTDHSDELNCSCDAGHOFCKNPKFCKPLFWVCDVNDGDNSEGGSCPAQTFRC 532  
QY 361 SNGKCLSKSQOQNGKDCDGDSDDEASCPKYNVVTCTKHTYRCINGLCLSKGNBECDEKED 420  
DB 533 SNGKCLSKSQOQNGKDCDGDSDDEASCPKYNVVTCTKHTYRCINGLCLSKGNBECDEKED 592  
QY 421 CSQSDSEKDCDCLGRFTROARVVGCTDADBEQWPMQVSLHALGQGHICGASLISPMNLV 480  
DB 593 CSQSDSEKDCDCLGRFTROARVVGCTDADBEQWPMQVSLHALGQGHICGASLISPMNLV 652  
QY 481 SAHHCYIDDRGFRYSPTQMTAFLGLHDOSQSAFQVQERRLRKRIISHFPFNDFTFYDI 540  
DB 653 SAHHCYIDDRGFRYSPTQMTAFLGLHDOSQSAFQVQERRLRKRIISHFPFNDFTFYDI 712  
QY 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTMGHTQYGGTGALLLOKGEIRVIN 600  
DB 713 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTMGHTQYGGTGALLLOKGEIRVIN 772  
QY 601 QTTCCENILLPQOITPRMNCVGFSLGGVDSCCGDSGGGLSSVEADGRIFQAGVWSGDCQAQ 660

## RESULT 2

US-09-644-600-2  
; Sequence 2, Application US/09644600  
; Patent No. 6431500  
; GENERAL INFORMATION:  
; APPLICANT: Tanimoto, Timothy J.  
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
; FILE REFERENCE: D6064CIP/D  
; CURRENT APPLICATION NUMBER: US/09/644,600  
; CURRENT FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: 09/421,213  
; PRIOR FILING DATE: 1999-10-20  
; PRIOR APPLICATION NUMBER: 09/027,337  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 2  
; LENGTH: 855  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: TADG-15  
US-09-644-600-2

Query Match 99.1%; Score 3754; DB 4; Length 855;  
Best Local Similarity 99.3%; Pred. No. 3.9e-262;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNSCSFGIHAAGVEIMRFTTPG 60  
DB 173 MAERVVMLEPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNSCSFGIHAAGVEIMRFTTPG 232  
QY 61 FPDSPYPAHARCOMALRGADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120  
DB 233 FPDSPYPAHARCOMALRGADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292  
QY 121 LCGTTPPSYNLTFHSSQNVLLITLINTERRRHGFEEATFFOLPRMSSCGRLRKAQGTEN 180  
DB 293 LCGTTPPSYNLTFHSSQNVLLITLINTERRRHGFEEATFFOLPRMSSCGRLRKAQGTEN 352  
QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVAVRPFKFFYLLEPRRACGTCPKDYVEINGEKYCGE 240  
DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVAVRPFKFFYLLEPRRACGTCPKDYVEINGEKYCGE 412  
QY 241 RSQFVVTSSNKTITVRFHSDQSYTDTGFLAELYSYSSDPCPGQFTCRGRICRKELRCD 300  
DB 413 RSQFVVTSSNKTITVRFHSDQSYTDTGFLAELYSYSSDPCPGQFTCRGRICRKELRCD 472  
QY 301 GMADCTDHSDELNCSCDAGHOFCKNPKFCKPLFWVCDVNDGDNSEGGSCPAQTFRC 360  
DB 473 GMADCTDHSDELNCSCDAGHOFCKNPKFCKPLFWVCDVNDGDNSEGGSCPAQTFRC 532  
QY 361 SNGKCLSKSQOQNGKDCDGDSDDEASCPKYNVVTCTKHTYRCINGLCLSKGNBECDEKED 420  
DB 533 SNGKCLSKSQOQNGKDCDGDSDDEASCPKYNVVTCTKHTYRCINGLCLSKGNBECDEKED 592  
QY 421 CSQSDSEKDCDCLGRFTROARVVGCTDADBEQWPMQVSLHALGQGHICGASLISPMNLV 480  
DB 593 CSQSDSEKDCDCLGRFTROARVVGCTDADBEQWPMQVSLHALGQGHICGASLISPMNLV 652  
QY 481 SAHHCYIDDRGFRYSPTQMTAFLGLHDOSQSAFQVQERRLRKRIISHFPFNDFTFYDI 540  
DB 653 SAHHCYIDDRGFRYSPTQMTAFLGLHDOSQSAFQVQERRLRKRIISHFPFNDFTFYDI 712

```
Qy 541 ALLELEKPAEYSSMWVRICLPDASHVPAGKAIWYTGHTGYGTCGALLILOKEIRVIN 600
Db 713 ALLELEKPAEYSSMWVRICLPDASHVPAGKAIWYTGHTGYGTCGALLILOKEIRVIN 772
Qy 601 QTTCEMLLPQOITPRMNCVGFSLSGVDSCCGDSGGLSSVEADGRIFQAGVWSMGDCQAQ 660
Db 773 QTTCEMLLPQOITPRMNCVGFSLSGVDSCCGDSGGLSSVEADGRIFQAGVWSMGDCQAQ 832
Qy 661 RNKPGVYTRLPFRDWMIKENTGV 683
Db 833 RNKPGVYTRLPFRDWMIKENTGV 855

RESULT 3
US-09-654-600A-2
; Sequence 2, Application US/09654600A
; Patent No. 6649741
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; 09/027,337
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-654-600A-2

Query Match 99.1%; Score 3754; DB 4; Length 855;
Best Local Similarity 99.3%; Pred. No. 3.9e-262;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAEERVWMLPPRARSLSKSFVVTSVVAPPTDSKTQVORTODNSCSFGLHARGVELMRFTTTPG 60
Db 173 MAEERVWMLPPRARSLSKSFVVTSVVAPPTDSKTQVORTODNSCSFGLHARGVELMRFTTTPG 232
Qy 61 PPDSPPYPAHARCOMALRGDADSVLSLTFRSPFDLASCDERSGDLVTVYNTLSPMEPHALVQ 120
Db 233 PPDSPPYPAHARCOMALRGDADSVLSLTFRSPFDLASCDERSGDLVTVYNTLSPMEPHALVQ 292
Qy 121 LCGTTPPSYNLTFFHSSQNVLLITLTNTERHHPGFATFPOLPRMSSCGGLRLKAQGTFN 180
Db 293 LCGTTPPSYNLTFFHSSQNVLLITLTNTERHHPGFATFPOLPRMSSCGGLRLKAQGTFN 352
Qy 181 SPYYPGHYPPNIDCTWNIIEVNNQHVKVRFPFYLLLEPRBACGTCPCPDYVEINGEKYCGE 240
Db 353 SPYYPGHYPPNIDCTWNIIEVNNQHVKVRFPFYLLLEPRBACGTCPCPDYVEINGEKYCGE 412
Qy 241 RSQFVVTSSNSNKTIVRFHSDOSYTDGTGLAEYLSYDSDPCPGQFTCTGRICRKELRCD 300
Db 413 RSQFVVTSSNSNKTIVRFHSDOSYTDGTGLAEYLSYDSDPCPGQFTCTGRICRKELRCD 472
Qy 301 GWADCTHSDDELNSCDAGHOFCTCKNFKCKPLFWVCDVNDGDNSSDEGSCSPAQTRFC 360
Db 473 GWADCTHSDDELNSCDAGHOFCTCKNFKCKPLFWVCDVNDGDNSSDEGSCSPAQTRFC 532
Qy 361 SNGKLSKSOQONGKDCGDSDEASCPKVVVVTCTKHTYRCINGLCLSKNPECDGKED 420
Db 533 SNGKLSKSOQONGKDCGDSDEASCPKVVVVTCTKHTYRCINGLCLSKNPECDGKED 592
Qy 421 CSDDSDDEKDCGCGRLSFTROARVVGCTDADSGEMFWQVSLHALGGHICGASLISPMNLV 480
Db 593 CSDDSDDEKDCGCGRLSFTROARVVGCTDADSGEMFWQVSLHALGGHICGASLISPMNLV 652
```

```
Qy 481 SAACHYIDBGFPRYSDEPTQMTAFGLHDOSQSRASAPGVQERLRKRIISHPFENFTPDYDI 540
Db 653 SAACHYIDBGFPRYSDEPTQMTAFGLHDOSQSRASAPGVQERLRKRIISHPFENFTPDYDI 712
Qy 541 ALLELEKPAEYSSMWVRICLPDASHVPAGKAIWYTGHTGYGTCGALLILOKEIRVIN 600
Db 713 ALLELEKPAEYSSMWVRICLPDASHVPAGKAIWYTGHTGYGTCGALLILOKEIRVIN 772
Qy 601 QTTCEMLLPQOITPRMNCVGFSLSGVDSCCGDSGGLSSVEADGRIFQAGVWSMGDCQAQ 660
Db 773 QTTCEMLLPQOITPRMNCVGFSLSGVDSCCGDSGGLSSVEADGRIFQAGVWSMGDCQAQ 832
Qy 661 RNKPGVYTRLPFRDWMIKENTGV 683
Db 833 RNKPGVYTRLPFRDWMIKENTGV 855

RESULT 4
US-09-644-600-10
; Sequence 10, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 10
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Epitchin
US-09-644-600-10

Query Match 82.4%; Score 3123; DB 4; Length 902;
Best Local Similarity 82.9%; Pred. No. 1.1e-216;
Matches 556; Conservative 56; Mismatches 59; Indels 0; Gaps 0;

Qy 1 MAEERVWMLPPRARSLSKSFVVTSVVAPPTDSKTQVORTODNSCSFGLHARGVELMRFTTTPG 60
Db 173 MAEERVWMLPPRARSLSKSFVVTSVVAPPTDSKTQVORTODNSCSFGLHARGVELMRFTTTPG 232
Qy 61 PPDSPPYPAHARCOMALRGDADSVLSLTFRSPFDLASCDERSGDLVTVYNTLSPMEPHALVQ 120
Db 233 PPDSPPYPAHARCOMALRGDADSVLSLTFRSPFDLASCDERSGDLVTVYNTLSPMEPHALVQ 292
Qy 121 LCGTTPPSYNLTFFHSSQNVLLITLTNTERHHPGFATFPOLPRMSSCGGLRLKAQGTFN 180
Db 293 LCGTTPPSYNLTFFHSSQNVLLITLTNTERHHPGFATFPOLPRMSSCGGLRLKAQGTFN 352
Qy 181 SPYYPGHYPPNIDCTWNIIEVNNQHVKVRFPFYLLLEPRBACGTCPCPDYVEINGEKYCGE 240
Db 353 SPYYPGHYPPNIDCTWNIIEVNNQHVKVRFPFYLLLEPRBACGTCPCPDYVEINGEKYCGE 412
Qy 241 RSQFVVTSSNSNKTIVRFHSDOSYTDGTGLAEYLSYDSDPCPGQFTCTGRICRKELRCD 300
Db 413 RSQFVVTSSNSNKTIVRFHSDOSYTDGTGLAEYLSYDSDPCPGQFTCTGRICRKELRCD 472
Qy 301 GWADCTHSDDELNSCDAGHOFCTCKNFKCKPLFWVCDVNDGDNSSDEGSCSPAQTRFC 360
Db 473 GWADCTHSDDELNSCDAGHOFCTCKNFKCKPLFWVCDVNDGDNSSDEGSCSPAQTRFC 532
Qy 361 SNGKLSKSOQONGKDCGDSDEASCPKVVVVTCTKHTYRCINGLCLSKNPECDGKED 420
Db 533 SNGKLSKSOQONGKDCGDSDEASCPKVVVVTCTKHTYRCINGLCLSKNPECDGKED 592
```

Db 533 SNGKCLPQSOXKNGKNDKCGSDSEASCDVNVVVSCTKYTRCQNGCLCLSKGNEPCDEKTD 592  
Qy 421 CSDSDSEKDCDCGLRSTFRQARVVGTTADGEMPMQVSLHALGQGHICGASLISPMNLV 480  
Db 593 CSDSDSEKDCDCGLRSTFRQARVVGTTADGEMPMQVSLHALGQGHICGASLISPMNLV 652  
Qy 481 SAAHCYIDRGRFYSPTQMTAFGLHDOSQASAPGVQERLKRILSHPFNDFTFDYDI 540  
Db 653 SAAHCYIDRGRFYSPTQMTAFGLHDOSQASAPGVQERLKRILSHPFNDFTFDYDI 712  
Qy 541 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWMTGHTQYGGTALLQKGEIRVIN 600  
Db 713 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWMTGHTGEGTALLQKGEIRVIN 772  
Qy 601 QTTCEMLPQOITPRMNCVGLSGVDSCQSDSGPLSVYADGRIFQAGVSWGDCAQ 660  
Db 773 QTTCEMLPQOITPRMNCVGLSGVDSCQSDSGPLSVYADGRIFQAGVSWGDCAQ 832  
Qy 661 RNKPGVYTRLP 671  
Db 833 RNKPGVYTRLP 843

RESULT 5  
US-09-654-600A-10  
; Sequence 10, Application US/09654600A  
; Patent No. 6649741  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Tanimoto, Hiroshi  
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
; FILE REFERENCE: D604CIP/D  
; CURRENT APPLICATION NUMBER: US/09/654,600A  
; CURRENT FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 09/421,213  
; 09/027,337  
; PRIOR FILING DATE: 1999-10-20  
; 1998-02-20  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 10  
; LENGTH: 902  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
; OTHER INFORMATION: Epithin  
US-09-654-600A-10

Query Match 82.4%; Score 3123; DB 4; Length 902;  
Best Local Similarity 82.9%; Pred. No. 1,1e-216;  
Matches 556; Conservative 56; Mismatches 59; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRAALSLKSVVTVSVAPPTDSKYQORQDSCSGHLAGVYELMRFTTPG 60  
Db 173 MAERVVMLPPRAALSLKSVVTVSVAPPTDSKYQORQDSCSGHLAGVYELMRFTTPG 232  
Qy 61 FPDSPYAHARCOVALRGDADSVLSTFRSFDLASCEBGSGLDLYTVNTLSPMEPHALVQ 120  
Db 233 FPDSPYAHARCOVALRGDADSVLSTFRSFDVASCEBGSGLDLYTVNTLSPMEPHALV 292  
Qy 121 LCGTPSPYNLTFRHSQNVLLITLITNTERRHGFQATFQOLPMSSCGGLRKAQGTN 180  
Db 293 LCGTPSPYNLTFRHSQNVLLITLITNTERRHGFQATFQOLPMSSCGGLRKAQGTN 352  
Qy 181 SPYRPGHYPRINDCTNMIIEVPPNNGHYVVRKFFFLLEPRACGCPDQVYEINEKXCGE 240  
Db 353 SPYRPGHYPRINDCTNMIIEVPPNNGHYVVRKFFFLLEPRACGCPDQVYEINEKXSGE 412  
Qy 241 RSQGVTVSNKTIIVRFHSDSYDTGFLAEYLSDSSDPCPGQFTCRGTGCIKELRCD 300  
Db 413 RSQGVTVSNKTIIVRFHSDSYDTGFLAEYLSDSSDPCPGQFTCRGTGCIKELRCD 472  
Qy 301 GWADCTDHSDELNCSGDAQHQTCKNRCXPLFWVCDSDVNDGSDSDGSCSCPAQTFR 360

Db 473 GWADCEPDYSBERYCRGNATHTQFTCKNQFCXPLFWVCDSDVNDGSDSDGSCSCPAQ 532  
Qy 361 SNGKCLSKSQOCCXKDDCGSDSEASCPYNNVYTCRKHYYRCNLNGCLSKGNEPCCKED 420  
Db 533 SNGKCLPQSOXKNGKNDKCGSDSEASCDVNVVVSCTKYTRCQNGCLCLSKGNEPCCKED 592  
Qy 421 CSDSDSEKDCDCGLRSTFRQARVVGTTADGEMPMQVSLHALGQGHICGASLISPMNLV 480  
Db 593 CSDSDSEKDCDCGLRSTFRQARVVGTTADGEMPMQVSLHALGQGHICGASLISPMNLV 652  
Qy 481 SAAHCYIDRGRFYSPTQMTAFGLHDOSQASAPGVQERLKRILSHPFNDFTFDYDI 540  
Db 653 SAAHCYIDRGRFYSPTQMTAFGLHDOSQASAPGVQERLKRILSHPFNDFTFDYDI 712  
Qy 541 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWMTGHTQYGGTALLQKGEIRVIN 600  
Db 713 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWMTGHTGEGTALLQKGEIRVIN 772  
Qy 601 QTTCEMLPQOITPRMNCVGLSGVDSCQSDSGPLSVYADGRIFQAGVSWGDCAQ 660  
Db 773 QTTCEMLPQOITPRMNCVGLSGVDSCQSDSGPLSVYADGRIFQAGVSWGDCAQ 832  
Qy 661 RNKPGVYTRLP 671  
Db 833 RNKPGVYTRLP 843

RESULT 6  
US-09-657-986B-2  
; Sequence 2, Application US/09657986B  
; Patent No. 6797504  
; GENERAL INFORMATION:  
; APPLICANT: Madison, Edwin L.  
; APPLICANT: Semple, Joseph Edward  
; APPLICANT: Coombs, Gary Samuel  
; APPLICANT: Reiner, John Eugene  
; APPLICANT: Ong, Edgar O.  
; APPLICANT: Arai, Gian Luca  
; TITLE OF INVENTION: Inhibitors of Serine Protease Activity of Matrilysin or  
; FILE REFERENCE: Corvas 255/049  
; CURRENT APPLICATION NUMBER: US/09/657,986B  
; CURRENT FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-09-657-986B-2

Query Match 34.8%; Score 1319; DB 4; Length 241;  
Best Local Similarity 100.0%; Pred. No. 1.8e-87;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 VVGSTADDEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDRGRFYSPTQMTA 502  
Db 1 VVGSTADDEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDRGRFYSPTQMTA 60  
Qy 503 FLGLHDOSQASAPGVQERLKRILSHPFNDFTFDYDIALLELEKPAEYSSMWRPCLPD 562  
Db 61 FLGLHDOSQASAPGVQERLKRILSHPFNDFTFDYDIALLELEKPAEYSSMWRPCLPD 120  
Qy 563 ASHVFPAGKAIWMTGHTQYGGTALLQKGEIRVINQTTCEMLPQOITPRMNCVGL 622  
Db 121 ASHVFPAGKAIWMTGHTQYGGTALLQKGEIRVINQTTCEMLPQOITPRMNCVGL 180  
Qy 623 SGGVDSGQSDSGPLSVYADGRIFQAGVSWGDCAQRNKPGVYTRLPFRDMIKENTG 682  
Db 181 SGGVDSGQSDSGPLSVYADGRIFQAGVSWGDCAQRNKPGVYTRLPFRDMIKENTG 240  
Qy 683 V 683

Db 241 V 241

RESULT 7  
US-08-200-900A-2

Sequence 2, Application US/08200900A

Patent No. 5655566

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/200,900A

FILING DATE: 23-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Meinhart, Maureen C.

REGISTRATION NUMBER: 31,544

REFERENCE/DOCKET NUMBER: GI 5201-FWC

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170 X8574

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 798 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-200-900A-2

Query Match 18.6%; Score 703.5; DB 1; Length 798;

Best Local Similarity 32.6%; Pred. No. 1.5e-42;

Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;

Qy 161 QLPMSGCG--RIRKAGTFSNPPYRGHYPNIDCTWNIEVNNQHVKRFKFFYLLEP 218  
Db 298 ELP--TDCGGPHDLMEBNTTFTSINFPNSYPNOAFCIWNLNAQKKNIQLFQEFDELN- 354  
Qy 219 RRACGTCPKDYVEI--NGEKYCGERSQFV-----VTSNKNITVRFHSDQSYTDT 266  
Db 355 -----IADVEIRDE--GDSLFVAVTGPVNDVSTTNRMTVLFITDNMLAKQ 404  
Qy 267 GFLLAEYLS--YDSSDPC-PGQFTCTGRCIRKELRCDGWADCTDHSDELNC-----SC 316  
Db 405 GFKANFTTGYGLGPECKEDNFQCKGECIPLVNLCDGPFHCKDGSDEAHCVRLFNGT 464  
Qy 317 DAGH--OPTCKNKKCKPLFWVCDSVNDGSDSGSCSPAQTRFGSNGKLSLSSQCGNG 374  
Db 465 DSSGLVQRRIS-----IMHV-----ACAEN-----WTTQISDVC-----QLLG 499  
Qy 375 KDPCDGSDEASG-----PRVNVVCTKHTYRCINGLCLSKGNPBCDKEDCSGSDSEK 428  
Db 500 ---LGTNNSVPTSTGGGPFVNLNTAP-----NSGLILTPSQGC-----LDSILIL 544  
Qy 429 DCD---CGLASFTQA--RVVIGTDADEGEMPMOVSLLHALGQGHICGASLSIPMWLYSAA 483  
Db 545 QCNYSKCGKLVLTQVSEPKIVGSGDSREGAMPVVALYFDDQ--QVCGASLVSRLMLYSAA 603  
Qy 484 HCYIDDBGFRSDPTQWTAFLGLHDGQSRASAPGVERTLKRILISHPFNDPTFDYDALL 543

Db 604 HCYVG-----NMESKWKAVLGLHMASNLTSPOJETRLIQIYIVNPHYKRRKNDIAM 659  
Qy 544 ELEKPAEYSMAVRPICLPDASHVFPACKAIWVGHTYGGGALLIQGEIRVINQTT 603  
Db 660 HLEKWKVYTDYIOPICLPEENOVFPFGRICSIAGWGLIYQGSTADVLQADAVPLLSNEK 719  
Qy 604 CENLLPQ-QITPRMVCVGLSGVDSCQSGSPPLSVADGRIFQAGVYVSWGDCQARN 662  
Db 720 CQQQMPFYNTENNVCAGYRAGVDSCQSGSGPLMCQF--NRMWLNGVTSFGYCALPN 778  
Qy 663 KPGVYTRLPLEFDMIX 678  
Db 779 RPYVYARVPRTFEMIQ 794

RESULT 8  
US-08-794-042-2

Sequence 2, Application US/08794042

Patent No. 6746859

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/794,042

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/200,900

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Meinhart, Maureen C.

REGISTRATION NUMBER: 31,544

REFERENCE/DOCKET NUMBER: GI 5201-FWC

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170 X8574

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 798 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-794-042-2

Query Match 18.6%; Score 703.5; DB 4; Length 798;

Best Local Similarity 32.6%; Pred. No. 1.5e-42;

Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;

Qy 161 QLPMSGCG--RIRKAGTFSNPPYRGHYPNIDCTWNIEVNNQHVKRFKFFYLLEP 218  
Db 298 ELP--TDCGGPHDLMEBNTTFTSINFPNSYPNOAFCIWNLNAQKKNIQLFQEFDELN- 354  
Qy 219 RRACGTCPKDYVEI--NGEKYCGERSQFV-----VTSNKNITVRFHSDQSYTDT 266  
Db 355 -----IADVEIRDE--GDSLFVAVTGPVNDVSTTNRMTVLFITDNMLAKQ 404  
Qy 267 GFLLAEYLS--YDSSDPC-PGQFTCTGRCIRKELRCDGWADCTDHSDELNC-----SC 316  
Db 405 GFKANFTTGYGLGPECKEDNFQCKGECIPLVNLCDGPFHCKDGSDEAHCVRLFNGT 464  
Qy 317 DAGH--OPTCKNKKCKPLFWVCDSVNDGSDSGSCSPAQTRFGSNGKLSLSSQCGNG 374  
Db 465 DSSGLVQRRIS-----IMHV-----ACAEN-----WTTQISDVC-----QLLG 499  
Qy 375 KDPCDGSDEASG-----PRVNVVCTKHTYRCINGLCLSKGNPBCDKEDCSGSDSEK 428  
Db 500 ---LGTNNSVPTSTGGGPFVNLNTAP-----NSGLILTPSQGC-----LDSILIL 544  
Qy 429 DCD---CGLASFTQA--RVVIGTDADEGEMPMOVSLLHALGQGHICGASLSIPMWLYSAA 483  
Db 545 QCNYSKCGKLVLTQVSEPKIVGSGDSREGAMPVVALYFDDQ--QVCGASLVSRLMLYSAA 603  
Qy 484 HCYIDDBGFRSDPTQWTAFLGLHDGQSRASAPGVERTLKRILISHPFNDPTFDYDALL 543

Dh 405 GFKANFTTGYGLIPEPCCKEDNFQCKDECIPLVNLCDGFPFHCXGSDGAHCVRLFNQTT 464  
Qy 317 DAGH--OFTCKNKECKEPLFWVCDVNDGDNDSDEGSCPAQTPRCNSGKCLSKSQOQNG 374  
Dh 465 DSSGLVQFRIQS-----IMHV-----ACAEN-----WTTQISDDVC-----QLLG 499  
Qy 375 KDCCGSDGSDASG-----PKVNVVCTKHTYRCLNGLCLSKGNPECDGKEDSDGDEK 428  
Dh 500 ---LGTGNSSVPRFTSTGGFPYVNLNTAP-----NSGLITPESQOC-----LEDSTILL 544  
Qy 429 DCD---CGLSFTROA--RVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLVSA 483  
Dh 545 QCNVKSGCKKLTVQEVSPKIVGSSDSRREGAMPVVALYFPDQ--QVCGASLVSRLVLSAA 603  
Qy 484 HCYTIDRGFRYSPTQWTAFLGLHDQSORAPGVQERLKRILSHPEFNDFTFYDIAL 543  
Dh 604 HCYYG-----NMESKMKAVLGLHMASNLTPQIETRLIDQIVNPNRRKNDIAM 659  
Qy 544 ELEKPAEYSSMVRPICLPDASHVFPAGKAIWVGHTQYGGTALILQKEIRVINQTT 603  
Dh 660 HLEKKNVYTYIOPICLPEENOVFPFGRICSIAGMGALIIYGSTADVLQEDAVPLLSNEK 719  
Qy 604 CENILPQ--QITPRMVCVFLSGVDSGQDGGPLSSVEADGRIFOAGVSWGDCGAQRN 662  
Dh 720 CQOQMPRYNITENNVCAGYEAGVDSGQDGGPLMCOE--NNRWLLAGVTSFGYQCALPN 778  
Qy 663 KPGVYTRPLPLFRDWIK 678  
Dh 779 RPGVYARVPRFTMIQ 794

RESULT 9  
PCT-US94-00616-2  
; Sequence 2, Application PC/TUS9400616  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE  
; NUMBER OF SEQUENCES: 33  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/00616  
; FILING DATE:  
; CLASSIFICATION:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 798 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US94-00616-2

Query Match 18.6%; Score 703.5; DB 5; Length 798;  
Best Local Similarity 32.6%; Pred. No. 1.5e-42;  
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;

Qy 161 QLRPMSSCGG--RLRKAQGFNPSYYGHPNIDCTWNIEVNNQHVAKRFKFFYLEP 218  
Dh 298 ELP--TDCGPHDLMENPTTFTSINFPNSYPNQAFCIMNIAOKGKNIQIHFQFIDLEN 354  
Qy 219 RBAAGCTPKQYVEI--NEKKYCGERSQPV-----VTSNKNKTIVAFHSHQSTDT 266  
Dh 355 -----IADVLRDGE-----GDSLPLAVYTGPGPNVDFSTNRRKTVLFTIDNMLAKQ 404  
Qy 267 GFLEAYLS---YDSSDPC--PQFTCRTRGRCIRKELRCDGMADCTDHSDELNC-----SC 316  
Dh 405 GFKANFTTGYGLIPEPCCKEDNFQCKDECIPLVNLCDGFPFHCXGSDGAHCVRLFNQTT 464  
Qy 317 DAGH--OFTCKNKECKEPLFWVCDVNDGDNDSDEGSCPAQTPRCNSGKCLSKSQOQNG 374

Dh 465 DSSGLVQFRIQS-----IMHV-----ACAEN-----WTTQISDDVC-----QLLG 499  
Qy 375 KDCCGSDGSDASG-----PKVNVVCTKHTYRCLNGLCLSKGNPECDGKEDSDGDEK 428  
Dh 500 ---LGTGNSSVPRFTSTGGFPYVNLNTAP-----NSGLITPESQOC-----LEDSTILL 544  
Qy 429 DCD---CGLSFTROA--RVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLVSA 483  
Dh 545 QCNVKSGCKKLTVQEVSPKIVGSSDSRREGAMPVVALYFPDQ--QVCGASLVSRLVLSAA 603  
Qy 484 HCYTIDRGFRYSPTQWTAFLGLHDQSORAPGVQERLKRILSHPEFNDFTFYDIAL 543  
Dh 604 HCYYG-----NMESKMKAVLGLHMASNLTPQIETRLIDQIVNPNRRKNDIAM 659  
Qy 544 ELEKPAEYSSMVRPICLPDASHVFPAGKAIWVGHTQYGGTALILQKEIRVINQTT 603  
Dh 660 HLEKKNVYTYIOPICLPEENOVFPFGRICSIAGMGALIIYGSTADVLQEDAVPLLSNEK 719  
Qy 604 CENILPQ--QITPRMVCVFLSGVDSGQDGGPLSSVEADGRIFOAGVSWGDCGAQRN 662  
Dh 720 CQOQMPRYNITENNVCAGYEAGVDSGQDGGPLMCOE--NNRWLLAGVTSFGYQCALPN 778  
Qy 663 KPGVYTRPLPLFRDWIK 678  
Dh 779 RPGVYARVPRFTMIQ 794

RESULT 10  
US-09-734-675-4  
; Sequence 4, Application US/09734675  
; Patent No. 6365391  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL000862  
; CURRENT APPLICATION NUMBER: US/09/734,675  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 407  
; TYPE: PRT  
; ORGANISM: Human  
; US-09-734-675-4

Query Match 15.5%; Score 588; DB 3; Length 407;  
Best Local Similarity 45.9%; Pred. No. 1.4e-34;  
Matches 119; Conservative 38; Mismatches 82; Indels 20; Gaps 6;

Qy 432 CGLR---SFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLVSAHCYTD 488  
Dh 162 CGTRRSKTLGQSLRIVGTEVEGEMPMQASLQWDG--SHRCGATLINATVLSAAHCFT-- 219  
Qy 489 DRGFRYSPTQWTAFLGLHDQSORAPGVQERLKRILSHPEFNDFTFYDIALLEKRP 548  
Dh 220 ---TKNPARWYASGCV-----TIKPSMKKGLRLIIVHEKXKHPSHDYDLSLAEISSP 270  
Qy 549 ABYSMVRPICLPDASHVFPAGKAIWVGHTQYGGTALILQKEIRVINQTTENIL 608  
Dh 271 VPTYNVAHVRCFLDASVEFQPGVMEVTFGALKNDGVSQNHRLRQAVTLIDATTQNE-- 328  
Qy 609 PQ-----QITPRMVCVFLSGVDSGQDGGPLSSVEADGRIFOAGVSWGDCGAQRN 664  
Dh 329 PQAYNDALITPRMVCASLSEKTDACQDGGPLVSSDARDIWTLAGIVSGDECAKPNRP 388  
Qy 665 GVTYTRPLPLFRDWIKENTGV 683  
Dh 389 GVTYTRVTLALRDWITTSKTI 407

```
RESULT 11
US-09-685-166A-895
; Sequence 895, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolck, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685.166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-895

Query Match          14.8%; Score 560.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 1.7e-32;
Matches 131; Conservative 58; Mismatches 128; Indels 69; Gaps 14;

QY 350 GCSCPAQTRC-SNGKCLSKSQCNKDKDCGDSDEASCPRV--NVV----- 393
DB 110 GSKCSNGIECDSSGTCINPNCDCGVSHCPGEBENRCVRLYGPNTLQMYSSQKSMH 169
QY 394 -----TCT-----KHTYRCLNG-----LCLSKNPECDKEDCSD 423
DB 170 PVCDDMNENYGRACRDMGYKNNFYSSQGIYVDSGSTSPMKLNTSAGNVDIYKKLYHSD 229
QY 424 GSDEK-----DCDCGLRSFTRQARVVGTDADGEPMQVSLHALGQGHICGASLIS 475
DB 230 ACSKAVAVSLRCLACGVNLNS-SRQSRIVGESALPGAMPQVSLH-VQNVHVCQGSIT 287
QY 476 PNMVLSAAHCYIDRGFRYSPTQWTAFLGLHDQ-SQSAFQVQERLKIISHPFENDF 534
DB 288 PEMVITVAHCVEKP---LNNPMWMTAFAGILRQSFMYGAGYQ---VQKVISHPNYDSK 340
QY 535 TFDVDIALLEKRAEYSWVRPCLPDASHVFPAGKAIWVTGNGHTQYGTGALLIQKG 594
DB 341 TKNDIDALMKLQKPLTFNDLVKPVCLPNPGMMLQPEOLCWSGATEBEKKTSEVLNAA 400
QY 595 EIRVINTTCEN--LLPQITPRMVCVFLSGVDSCCGDSGGPLSVLEADGRI-FQAGV 651
DB 401 KVLIIETORCNSRYVYDNLITPAMICAGFLQGNVDSCCGDSGGPL--VTSNNNIWMLIGD 458
QY 652 VSWGDCGAQRNKKPGVYTRLPFRDWI 677
DB 459 TSMGSCAKAYRPGVYGWVFTDWI 484

RESULT 12
US-09-879-792-14
; Sequence 14, Application US/09879792
; Patent No. 6734006
; GENERAL INFORMATION:
```

```
APPLICANT: Xiao, Yonghong
APPLICANT: Gedrich, Richard
; TITLE OF INVENTION: Regulation of Human Transmembrane Sertine
; FILE REFERENCE: 02973.00035
; CURRENT APPLICATION NUMBER: US/09/879.792
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/211,224
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: US 60/283,353
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/283,648
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: PCT
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-879-792-14

Query Match          14.8%; Score 560.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 1.7e-32;
Matches 131; Conservative 58; Mismatches 128; Indels 69; Gaps 14;

QY 350 GCSCPAQTRC-SNGKCLSKSQCNKDKDCGDSDEASCPRV--NVV----- 393
DB 110 GSKCSNGIECDSSGTCINPNCDCGVSHCPGEBENRCVRLYGPNTLQMYSSQKSMH 169
QY 394 -----TCT-----KHTYRCLNG-----LCLSKNPECDKEDCSD 423
DB 170 PVCDDMNENYGRACRDMGYKNNFYSSQGIYVDSGSTSPMKLNTSAGNVDIYKKLYHSD 229
QY 424 GSDEK-----DCDCGLRSFTRQARVVGTDADGEPMQVSLHALGQGHICGASLIS 475
DB 230 ACSKAVAVSLRCLACGVNLNS-SRQSRIVGESALPGAMPQVSLH-VQNVHVCQGSIT 287
QY 476 PNMVLSAAHCYIDRGFRYSPTQWTAFLGLHDQ-SQSAFQVQERLKIISHPFENDF 534
DB 288 PEMVITVAHCVEKP---LNNPMWMTAFAGILRQSFMYGAGYQ---VQKVISHPNYDSK 340
QY 535 TFDVDIALLEKRAEYSWVRPCLPDASHVFPAGKAIWVTGNGHTQYGTGALLIQKG 594
DB 341 TKNDIDALMKLQKPLTFNDLVKPVCLPNPGMMLQPEOLCWSGATEBEKKTSEVLNAA 400
QY 595 EIRVINTTCEN--LLPQITPRMVCVFLSGVDSCCGDSGGPLSVLEADGRI-FQAGV 651
DB 401 KVLIIETORCNSRYVYDNLITPAMICAGFLQGNVDSCCGDSGGPL--VTSNNNIWMLIGD 458
QY 652 VSWGDCGAQRNKKPGVYTRLPFRDWI 677
DB 459 TSMGSCAKAYRPGVYGWVFTDWI 484

RESULT 13
US-09-679-426-895
; Sequence 895, Application US/09679426
; Patent No. 6759515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolck, John A.
; APPLICANT: Vedvick, Thomas S.
```



```

; APPLICANT: Carter, Derrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C20
; CURRENT APPLICATION NUMBER: US/09/679.426
; CURRENT FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 895
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-679-426-895

Query Match      14.8%; Score 560.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 1.7e-32;
Matches 131; Conservative 58; Mismatches 128; Indels 69; Gaps 14;

QY 350 GCSCPAQTFRG-SNGKCLSKSQCNKGDDCGSDSDASCPKV---NVV----- 393
DB 110 GSKCSNGIGIECDSSGTCINPSNMCDGVSHCPGSEENRCVRLGPPFTLLQVYSSQKRSWH 169
QY 394 -----TCT---KHTYRCING-----LCISKGNPECDGKEDCSD 423
DB 170 PVCQDDMNENYGRAACRDMDGKNNFYSSQGIYDSDSGSTFMKLTNSAGNVDIYKKLYHSD 229
QY 424 GSDEK-----DCDGLRSFTTQARVVGCTDADGEMPMQVSLHALGQCHTIGASLIS 475
DB 230 ACSSKAVSLRCLACGNVLNS-SRQSRIVGGSALPGAMPQVSLH-VQNVHVCGSIIIT 287
QY 476 PNMLVSAHGYIDRGRFYSPTQMTAFGLGHDS-QRSAPGVQERLRKRIISHPFNDF 534
DB 288 PEMIVTAAHCEKRP---LNNPMTAFAGILRSFMYGAGVQ---VQKISHPNYDSK 340
QY 535 TFDYDIALLELEKPAEYSSMVRPCLPDASHVPAGKAIWVTGHTQYGGTALILQKG 594
DB 341 TKNNDIAMKQLQKPLFNDLVKVCPLPNRGMMLQPEQLCMISGMGATEBKGTSEVLNAA 400
QY 595 EIRVINGTTCEN--LLPQOITPRMCMCVGLSGGVDSQCGDSGGPLSSVEADGRI-FOAGV 651
DB 401 KVLIIETQRCSRYVDNLITTPAMICAGFLQGNVDSQCGDSGGPL--VTSNNIMWLIGD 458
QY 652 VSMGDCGAQRKRPVYTRLPFRDVI 677
DB 459 TSMGSGCAKAYRPGVYGNVWFDTWI 484

RESULT 14
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342.749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091.044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2
```

```

Query Match      14.7%; Score 558.5; DB 3; Length 492;
Best Local Similarity 33.9%; Pred. No. 2.3e-32;
Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;

QY 350 GCSCPAQTFRG-SNGKCLSKSQCNKGDDCGSDSDASCPKV---NVV----- 393
DB 110 GSKCSNGIGIECDSSGTCINPSNMCDGVSHCPGSEENRCVRLGPPFTLLQVYSSQKRSWH 169
QY 394 -----TCT---KHTYRCING-----LCISKGNPECDGKEDCSD 423
DB 170 PVCQDDMNENYGRAACRDMDGKNNFYSSQGIYDSDSGSTFMKLTNSAGNVDIYKKLYHSD 229
QY 424 GSDEK-----DCDGLRSFTTQARVVGCTDADGEMPMQVSLHALGQCHTIGASLIS 475
DB 230 ACSSKAVSLRCLACGNVLNS-SRQSRIVGGSALPGAMPQVSLH-VQNVHVCGSIIIT 287
QY 476 PNMLVSAHGYIDRGRFYSPTQMTAFGLGHDS-QRSAPGVQERLRKRIISHPFNDF 534
DB 288 PEMIVTAAHCEKRP---LNNPMTAFAGILRSFMYGAGVQ---VEKVISHPNYDSK 340
QY 535 TFDYDIALLELEKPAEYSSMVRPCLPDASHVPAGKAIWVTGHTQYGGTALILQKG 594
DB 341 TKNNDIAMKQLQKPLFNDLVKVCPLPNRGMMLQPEQLCMISGMGATEBKGTSEVLNAA 400
QY 595 EIRVINGTTCEN--LLPQOITPRMCMCVGLSGGVDSQCGDSGGPLSSVEADGRI-FOAGV 651
DB 401 KVLIIETQRCSRYVDNLITTPAMICAGFLQGNVDSQCGDSGGPL--VTSKNNIMWLIGD 458
QY 652 VSMGDCGAQRKRPVYTRLPFRDVI 677
DB 459 TSMGSGCAKAYRPGVYGNVWFDTWI 484

RESULT 15
US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691.840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342.749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091.044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-691-840-2

Query Match      14.7%; Score 558.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 2.3e-32;
Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;

QY 350 GCSCPAQTFRG-SNGKCLSKSQCNKGDDCGSDSDASCPKV---NVV----- 393
DB 110 GSKCSNGIGIECDSSGTCINPSNMCDGVSHCPGSEENRCVRLGPPFTLLQVYSSQKRSWH 169
QY 394 -----TCT---KHTYRCING-----LCISKGNPECDGKEDCSD 423
DB 170 PVCQDDMNENYGRAACRDMDGKNNFYSSQGIYDSDSGSTFMKLTNSAGNVDIYKKLYHSD 229
QY 424 GSDEK-----DCDGLRSFTTQARVVGCTDADGEMPMQVSLHALGQCHTIGASLIS 475
```

Db 230 ACSSKAVVSLRCLACGVNLNS-SRQSRIVGESALPGAMPQVSLH-VQNVHCGGSIIT 287  
Qy 476 PNMVLSAHACYIDRGRFYSDPTQWTAFLGLHDOS-QRSAPGYVERRLKRIISHPFENDF 534  
Db 288 PEMVLTAAHCEKPE---LNNPMHTAPAGILRQSFMYGAGYQ---VEKVISHPNDYSK 340  
Qy 535 TFDVDIALLELEPAEYSSMWRLPICLPDASHVFPAGKAIWYTGWHTOYGGTGALLIQKG 594  
Db 341 TKNDNIALMKQKXPLTFNDLVKPYCLPMPGMLQPEQLCWTISGNGATEKGTSEVLNAA 400  
Qy 595 EIRVINTTCEN--LLPQOITPRMVCVGLSGVDSCGDSGGPLSSVEADGRI-FOAGV 651  
Db 401 KVLLEIQRCSRYRYDDNLITPAMICAGFLQGNVSDCGDSGGPL--VTSKNINIMWLIGD 458  
Qy 652 VSMGDGCAQRNKPQVYTRPLPLFRDWI 677  
Db 459 TSMGSGCAKAYRPGVYGNVWFTDWT 484

## RESULT 16

US-08-944-483-65  
Sequence 65, Application US/08944483  
Patent No. 6232456  
GENERAL INFORMATION:  
APPLICANT: COHEN, MAURICE  
APPLICANT: COLPITTS, TRACEY L.  
APPLICANT: FRIEDMAN, PAULA N.  
APPLICANT: GRANADOS, EDWARD N.  
APPLICANT: KLAS, MICHAEL R.  
APPLICANT: RUSSELL, JOHN C.  
APPLICANT: STEWART, KENT D.  
APPLICANT: STROUPE, STEVEN D.  
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
NUMBER OF INVENTION: OF THE PROSTATE  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,483  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6183-US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 6232456e  
US-08-944-483-65  
Query Match 14.4%; Score 547.5; DB 3; Length 235;

Best Local Similarity 44.7%; Pred. No. 6.1e-32;  
Matches 106; Conservative 44; Mismatches 80; Indels 7; Gaps 4;  
Qy 443 VVGSTDADEGEWMPQVSLHALGQGHICGASLISPNMVLVSAHACYIDRGRFYSDPTQWTA 502  
Db 1 IVGGSNAKEGAMPVNVLY-YGGRLLCGASLVSSDMLVSAHCVYQ---NNLEPSKMTA 55  
Qy 503 FLGLHDOSRSPAPGVQRRRLKRIISHPFNDFTFDYDIALLELEKPAEYSSMWRLPICLPD 562  
Db 56 ILGLHMSNLTSPTQVTRLLDEIVINPHYRRRDNDIAMHLEFKVNTDYIQLCLPE 115  
Qy 563 ASHVFPAGKAIWYTGWHTOYGGTGALLIQKEIRVINTTCENLPO-QITPRMVCVGF 621  
Db 116 ENQVFPGRNCISLIGMGTVVYQGTANILQEADVPLLSNRCQOQMPDEVYITERMICAGY 175  
Qy 622 LSGVDSCGDSGGPLSSVEADGRIFOAGVSMGDGCAQRNKPQVYTRPLPLFRDWIK 678  
Db 176 EEGIDSCQDSDGGPLMCQE-NNRWFLAGVTSFGYKCALPVRPGVAVRVSFTFMTQ 231

## RESULT 17

US-08-807-151-3  
Sequence 3, Application US/08807151  
Patent No. 6043033  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED  
TITLE OF INVENTION: PROTEASE  
NUMBER OF INVENTION: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/807,151  
FILING DATE: Filed Herewith  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PP-0227 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 416132  
US-08-807-151-3  
Query Match 14.4%; Score 544.5; DB 3; Length 235;  
Best Local Similarity 43.5%; Pred. No. 1e-31;  
Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;  
Qy 443 VVGSTDADEGEWMPQVSLHALGQGHICGASLISPNMVLVSAHACYIDRGRFYSDPTQWTA 502  
Db 1 IVGGSNAKEGAMPVNVLY-YGGRLLCGASLVSSDMLVSAHCVYQ---NNLEPSKMTA 55

Db 1 IVGSDSREGAMPVVALYFDDQ-QVCGASLVSRLDMLVSAHCYVG---NNMPSKMKKA 55  
Qy 503 FLGHDSORSAPGVOERRLKRIISHPFNDFTFYDIALLEKPAVSSMVPICLPD 562  
Db 56 VLGHMAASNLTSPOIEFRLDIQIYINPHYNRRKNDIAMHLEMKNYTYIPICLPE 115  
Qy 563 ASHVPFPAKAIWYTGWGHYOGGTGALLQKEIRVINOTTGENTLPO-QITPRMVCVF 621  
Db 116 ENQVFPFGRICSIAGMGALIIYQGSTADVLQADAVPLLSNEKCOQOMPEYNTENMVCAGY 175  
Qy 622 LSGVDSCQDSDSGPPLSSVADGRIFQAGVNSMGDGCQANKKGVYTRLPFRDWIK 678  
Db 176 EAGVDSCQDSDSGPPLMCOE-NNRWLLAGVTSFGYQCALPFRBGVYARVFRFTWIO 231  
RESULT 18  
; Sequence 3, Application US/09478957  
; Patent No. 6350448  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Lal, Preeti  
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED  
; TITLE OF INVENTION: PROTEASE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION NUMBER: US/09/478,957  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/807,151  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PR-0227 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 235 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Genbank  
; CLONE: 416132  
; US-09-478-957-3  
Query Match 14.4%; Score 544.5; DB 3; Length 235;  
Best Local Similarity 43.5%; Pred. No. 1e-31;  
Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

Qy 563 ASHVPFPAKAIWYTGWGHYOGGTGALLQKEIRVINOTTGENTLPO-QITPRMVCVF 621  
Db 116 ENQVFPFGRICSIAGMGALIIYQGSTADVLQADAVPLLSNEKCOQOMPEYNTENMVCAGY 175  
Qy 622 LSGVDSCQDSDSGPPLSSVADGRIFQAGVNSMGDGCQANKKGVYTRLPFRDWIK 678  
Db 176 EAGVDSCQDSDSGPPLMCOE-NNRWLLAGVTSFGYQCALPFRBGVYARVFRFTWIO 231  
RESULT 19  
; Sequence 2, Application US/09518046  
; Patent No. 6294663  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Underwood, Lowell J.  
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed  
; TITLE OF INVENTION: In Ovarian Carcinoma and Uses Thereof  
; FILE REFERENCE: D6192CIP  
; CURRENT APPLICATION NUMBER: US/09/518,046  
; CURRENT FILING DATE: 2000-03-02  
; EARLIER APPLICATION NUMBER: 09/261,416  
; EARLIER FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 153  
; SEQ ID NO 2  
; LENGTH: 454  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; OTHER INFORMATION: complete amino acid sequence of TMDG-12  
; OTHER INFORMATION: protein  
; US-09-518-046-2  
Query Match 14.3%; Score 543.5; DB 3; Length 454;  
Best Local Similarity 33.6%; Pred. No. 2.5e-31;  
Matches 128; Conservative 54; Mismatches 128; Indels 71; Gaps 14;  
Qy 358 FRC-SNGKCLSKSQCGKXGKDCGSDGSDASCPRY-----NVVCTKHTYRC----- 402  
Db 78 YRCRSFKEIETTRCDGVSDCKGDDEYCVAVGGQNAVLYFTASWTKMCSDDWKGH 137  
Qy 403 -LNGCLSKGNP-----ECDG-----KEDCSD 423  
Db 138 YANVACQQLGFPSPVSSDNLVSLSGQFREFFVSIDHLLPDDKVTALHHSVYVRGCAS 197  
Qy 424 GS-DEKQCD-CGRSFTROARVYGGTDADGEMPMOVSILHAGHICGASLSPMWLV 481  
Db 198 GHVYTLQCTACGHR-GRYSRTVGNMSSLISQPMQASLQFGY-HLCGGSVITPLMIT 255  
Qy 482 AAHCYIDDRGFRYSDPTOWTAFGLHDSORSAPGVOERRLKRIISHPFNDFTFYDIA 541  
Db 256 AAHCYVD-----LYLPKSWTIQGLVSLDNPAP---SHLVKTIYVHSKYPKRLGNDIA 307  
Qy 542 LLELEKPAEYSSNVRPILCPDASHVFPAGKAIWYTGWGHYOGGTGALLQKEIRVINQ 601  
Db 308 LMKIAGEPLTFENMIQVCLPNSSENPDEKVCWTSAGATEDGDASPVILNHAVALISN 367  
Qy 602 TTC--ENILPQOITPRMVCVFLSGVDSCQDSDSGPPLSSVADGRIFQ-AGVNSMGDC 658  
Db 368 KICNHRDVGIGTISPSMLCAGYLTGVDSCQDSDSGPL--VCEERLMLKLVGTSFGIGC 425  
Qy 659 AQRNKGVYTRLPFRDWIK 679  
Db 426 AEVNRKGVYTRVTSFLDWIHE 446  
RESULT 20  
; Sequence 4, Application US/10177661  
; Patent No. 6794173  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, Dirk M.

APPLICANT: Virca, G. Duke  
TITLE OF INVENTION: DENDRITIC CELL TRANSMEMBRANE SERINE PROTEASE  
FILE REFERENCE: 3256-A  
CURRENT APPLICATION NUMBER: US/10/177,661  
CURRENT FILING DATE: 2002-06-20  
PRIOR APPLICATION NUMBER: US 60/299,606  
PRIOR FILING DATE: 2001-06-20  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 4  
LENGTH: 446  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-177-661-4

Query Match 13.9%; Score 527; DB 4; Length 446;  
Best Local Similarity 34.5%; Pred. No. 3.8e-30;  
Matches 129; Conservative 46; Mismatches 132; Indels 67; Gaps 14;

QY 333 FMVCDSDNDGSDNDEGGSCPAQTFRCNSNCKLSKSGCGKXDCGDSDEASCPK--- 389  
DB 98 FM-----OGHTGIRYKEQRESCPKHVR-----CDGVVDCKLKSDCLCVRPDW 141  
QY 390 -----VNVVTCIKHTYRCINGLCLSKNPECDKEDCSGSDKDC-----DCGL 434  
DB 142 DKSLLKTIYSSSHQWL---PICSSNMNDSEYSEKTCQQLGFRSRECPGSRVYISLGCHGL 198  
QY 435 RSFTROARVVGCTDADGEMPMQVSLHALGCGHICGASLISPNMLVSAHACYIDR--- 490  
DB 199 RAMT--RRIYAGLASDSKPMQVSLH-FGTHICGGLIDAQVLTAAHCFEYTRKVL 255  
QY 491 -GFRYSDPTQWTAFLG---LHDSORSAIPGQERLRKRISHPFNFTPTFDYALALE 546  
DB 256 EG-----WKVAGSINLHQLPEAS-----LAEITINSYTBEDDYDALMKLS 300  
QY 547 KPAEYSSMVRPICLPDASHVFPAGKAIWVTGHT-OYGTGALILKGEIRVINTTCE 605  
DB 301 KPLTSLAHHPACLPNMGQFTSLNETCMITGFGKTRRDDTSPLFLAEVQNLIDFKKN 360  
QY 606 NLL--PQGITPRMVCVGLSGGVDSGCGSGPLSSVEADGRITQAGVWSKGDCAQRNK 663  
DB 361 DYLVYDSYLTFRMMACADLRGRDSCGDSGGL-VEQNNRWYLAGVTSWGTGCGQRNK 419  
QY 664 PGVYTRLPFRDWI 677  
DB 420 PGVYTKVTEVLPWI 433

RESULT 21  
US-08-681-151-3  
Sequence 3, Application US/08681151  
Patent No. 5869637  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Bandman, Olga  
APPLICANT: Braxton, Scott Michael  
APPLICANT: Goll, Surya  
TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/681,151

FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0074US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 638 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GENBANK  
CLONE: 205011  
US-08-681-151-3

Query Match 13.9%; Score 526.5; DB 2; Length 638;  
Best Local Similarity 27.4%; Pred. No. 6.3e-30;  
Matches 184; Conservative 85; Mismatches 232; Indels 171; Gaps 32;

QY 99 RGSDLVTVVNTLSMEPHALVOLCGTTPSYNLTFFSSQNVLLITLTNTERHGF--E 156  
DB 32 RGGDLAIV---TPDAHC-QKMTFHPRLCLFSF-----LAVSPKEDDKRGCEWKE 81  
QY 157 ATFFQLPRMSSCG---GLRKAQGTNSPYYPGHPY-----NIDCTWNE-----V 200  
DB 82 SILETTERIRHRTGASISHSKCGHQLSACHODIYEGIDMRGSFNISKYDSIECOKLC 141  
QY 201 PNNOHV---VRFEPFLLEPRRAC-----GT-----C 225  
DB 142 TNNHCOFFYYATAFRRPFRYRSCCLKRSSSGTPIYSIKVDNLVSGFSLKSLSEIGC 201  
QY 226 PKD-----VYEINGEKYCGERS---QVVTSSNKITVRFHSDOSTYDGFALAYLSYD 276  
DB 202 PMDFQHFARADLNVSQVTPDAFVCRVCTFHPNCLFFFTYNEMWESQRYVCLKTS 261  
QY 277 SSD-PCP-----GQFTCRTRG--CIRKELRCGMDACTDHSDELNCSDAQH- 320  
DB 262 KSGRPSPIIOENAVSGYSLFTCRKARPERCHFKI-----YSGVAFGEELNATFVGAD 316  
QY 321 --QFTCKNRCKPLFWVCDSDVNDGSDSCPAQTFRCNSNCKLSKSGCGKXDC 378  
DB 317 ACQETCTKTRCOFFYVSLIPDC---KAEGCKC---SLRLST----- 353  
QY 379 GDGSDASCPKVVNVVTCIKHTYRCINGLCLSKNPECDKEDCSGSDKDC-----DC 432  
DB 354 -DSSP-----TRITY-----EAGS---SSGSLALCKVYESSDC 383  
QY 433 GLNSFTROARVVGCTDADGEMPMQVSLH--LGCGHICGASLISPNMLVSAHACYIDR 490  
DB 384 TTKI---NARIVGTSINSLGEMPMQVSLQVKLVSQNMHCGSIIIGRMWILTAHCF--D 437  
QY 491 GFRYSDPTQWTAFLGLHDSO--RSAPGQERLRKRISHPFNFTPTFDYALALEXP 548  
DB 438 GIPY--PDVWRITYGIIINLSIETNKP--FSSIKELIHOKTMSGSGYDILIKQTP 492  
QY 549 AEYSSMVRPICLPDASHVFPAGKAIWVTGHTOYGTGALILKGEIRVINTTCEML 608  
DB 493 LNTERPQKPICLPBKAADNTNITNCWTGWTGKREFTQNIQKATIPVPRBEQCKY 552  
QY 609 PQG-ITPRMVCVGLSGGVDSGCGSGPLSSVEADGRITQAGVWSKGDCAQRNKGY 667  
DB 553 RDVYITKQMICAGYKEGIDACKDSDGGL-VCKHSGRMQVLGITSGEGCARKKEQGVY 611

QY 668 TRLEPFDWKE 679  
Db 612 TKVAEYIDWLE 623

## RESULT 22

US-10-177-661-6  
; Sequence 6, Application US/10177661  
; Patent No. 6794173  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, Dick M.  
; TITLE OF INVENTION: DENDRITIC CELL TRANSMEMBRANE SERINE PROTEASE  
; FILE REFERENCE: 3256-A  
; CURRENT APPLICATION NUMBER: US/10/177,661  
; CURRENT FILING DATE: 2002-06-20  
; PRIOR APPLICATION NUMBER: US 60/299,606  
; PRIOR FILING DATE: 2001-06-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 418  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus sequence  
; NAME/KEY: MISC FEATURE  
; LOCATION: (1)-(418)  
; OTHER INFORMATION: Xaa = unknown  
US-10-177-661-6

Query Match 13.8%; Score 521.5; DB 4; Length 418;  
Best Local Similarity 30.9%; Pred. No. 8.9e-30;  
Matches 129; Conservative 43; Mismatches 160; Indels 85; Gaps 13;

QY 329 CKPLFWCDSDVNDGSDSCDSCGSCPAQTFRCNSNGKCLSKSCGCGKDKDCD 380  
Db 2 CIPKSMWCDGVXDCPKGEDEKXCVXXXXXGXXXXVRLYGDXLQVYSSSSXXXXPVCSD 61  
QY 381 GSDEA-----SCPKVNV-----VT 394  
Db 62 NMNBSYXXACXQWGXSAATYXSEKXXXXXGANSFKKLVNFPNNLXXDXYTSXIOXL 121  
QY 395 CTKHYRCLNGCLCLSKNPECDGKEDCSDSDKDCDCLRSFTROARVVGTDADGEW 454  
Db 122 XRSSXXCPSGXVSL-----QCS-----XQDCGVRLNAXXMTSRITGXXXXXGXW 168  
QY 455 PMQVSLHALGCGHICGASLISPMVLVSAHCVIIDDGRFYSDDP-TOMTAFLGLHDSQRS 513  
Db 169 PMQVSLQXXXXGVHLCCGSLIXPKVAVLTAHCVXG---RXXKPLXGKXVFAGILITXSLX 224  
QY 514 APQVQERRLKRIISHPFNDFTF-----DYDIALLEKPAEYSSMVRPICLPDASHVF 567  
Db 225 XPKAXXKXVKEKIIHFNYYXXXXXXSKNDIALMLSLKPLTFXYDIOVCLPDPQXL 284  
QY 568 PAKKAIWVTGHT-QYGGTGALLQKGEIRVINQTTCE--LLPQOITPRMVCVGLSG 624  
Db 285 XPGTTCWIXGWTATXKXGKTSPLQDAVPLIDNKKCNSYXVVDXKITPRMIGAGYLEG 344  
QY 625 GVDSCQDSGGLP---SSVEADGRIFOAGVVSWDGC-AQNRKPGVYTRLPFLFDWI 677  
Db 345 GVDSCQDSGGLPVCCEXXXXXQNNRMWLLXGXTSWGCGXAKANRPGVYTVYXFLXWI 401

## RESULT 23

US-10-177-661-2  
; Sequence 2, Application US/10177661  
; Patent No. 6794173  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, Dick M.  
; APPLICANT: Virca, G. Duke  
; TITLE OF INVENTION: DENDRITIC CELL TRANSMEMBRANE SERINE PROTEASE

; FILE REFERENCE: 3256-A  
; CURRENT APPLICATION NUMBER: US/10/177,661  
; CURRENT FILING DATE: 2002-06-20  
; PRIOR APPLICATION NUMBER: US 60/299,606  
; PRIOR FILING DATE: 2001-06-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 477  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-177-661-2

Query Match 13.7%; Score 518.5; DB 4; Length 477;  
Best Local Similarity 32.5%; Pred. No. 1.7e-29;  
Matches 135; Conservative 50; Mismatches 137; Indels 93; Gaps 17;

QY 292 CIRKEALRCQGMADCTSHSELNC-----SCDAGHOTCKNKKFKPLFW--VCD 337  
Db 114 CPGHAAVRCGVNDCKKSDBLGCVRPDWDKSLIKIYSSSHQ-----WLPIC 161  
QY 338 SVNDGSDSDSCGSCPAQTFRCNSNGKCLSKSQCGKDKDCDSDASCPKVVVYCTK 397  
Db 162 S--NMNDSYSEKTC-----QULGFESAHTTIVAHKDFANSFSLR 200  
QY 398 HTYRCLNGCLCLSKNPECDGKE---DCSDSDKDCDCLRSFTROARVVGTDADGE 453  
Db 201 YNSTIGESLHRS---ECPGQRYISLQCS-----HGLRAMT--GRIVGALASDSK 246  
QY 454 WPMQVSLHALGCGHICGASLISPMVLVSAHCVIIDDGRFYSDDP---GFRSDPTQWTAFIG--- 505  
Db 247 WPMQVSLH-FGTHICGGLIDAQWVLTAAHCFVTRKYLEG-----MNVVAGTSN 297  
QY 506 LHPQSORSAQVQERRLKRIISHPFNDFTFDYDIALLEKPAEYSSMVRPICLPDASH 565  
Db 298 LHPQPEAAS-----IAEIIINSYNTDEBDYDIALMLSKPLVLSAHIHAPCLPMHQ 350  
QY 566 VPAPKAIWVTGHT-QYGGTGALLQKGEIRVINQTTCE--LLPQOITPRMVCVGL 622  
Db 351 TFSINETCWITGFKRETDDKTSPLREYQVVLIFPKCNDVLYVDSYLTTPRMCGAGDL 410  
QY 623 SGVDSQDSGGLPLSSVEADGRIFOAGVVSWDGCAGNRKPGVYTRLPFLFDWI 677  
Db 411 RGRDSCQDSGGLP-VCEQNNRMWLLAGVTSWGTGCGQRNKPVGYYTKVTEVLPFI 464

## RESULT 24

US-09-879-792-12  
; Sequence 12, Application US/09879792  
; Patent No. 6734006  
; GENERAL INFORMATION:  
; APPLICANT: Xiao, Yonghong  
; APPLICANT: Gedrich, Richard  
; TITLE OF INVENTION: Regulation of Human Transmembrane Serine  
; FILE REFERENCE: 02973.00035  
; CURRENT APPLICATION NUMBER: US/09/879,792  
; CURRENT FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/211,224  
; PRIOR FILING DATE: 2000-06-13  
; PRIOR APPLICATION NUMBER: US 60/283,353  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: US 60/283,648  
; PRIOR FILING DATE: 2001-04-16  
; PRIOR APPLICATION NUMBER: PCT  
; PRIOR FILING DATE: 2001-06-12  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-879-792-12 (Docket No. 6734006 LIO-81-WO)



STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SCORN0701  
CLONE: 556016  
US-09-478-957-1

Query Match 13.5%; Score 512.5; DB 3; Length 283;  
Best Local Similarity 39.5%; Pred. No. 2.5e-29;  
Matches 111; Conservative 48; Mismatches 99; Indels 23; Gaps 9;

QY 409 SKNPEDCGKEDCDSDGDEK-----KDCDCGLRSFT-RQARVVGTDADGEGMPQVSL 460  
DB 6 SAGVVDLYKKLYHSDACSSRAVVALRCIACGVNINS--SRQSRIVGSGSALPGAMPQVSL 64  
QY 461 HALGQGHICGASLISPNMLVSAHACYIDDRGFYSDPTQMTAFGLHDG--QREAPVQVE 519  
DB 65 H-VQNVAVCCGSITTPETVITNAHCVEKP---LNNPWHMTAFGLRQSMFPGAGYO- 118  
QY 520 RLKRIISHPFNDPTDYDIALLEKPAEYSMWVPICLPDASHVFPAGKAIWVGW 579  
DB 119 --VEKVISHPNYDSKTNNDIALMKLQKPLFNDLVKPVCLPNEGMLQPEQLCMISGW 176  
QY 580 HTQYGGTALILQGEIRVINQTTCCEN--LIPQOITPRMVCVGLSGVDSQCGDSGGL 637  
DB 177 ATEKGGTSEVLNAKAVLLETQRCNSRYVDNLITPAMICAGFLQGVNDSQCGDSGGL 236  
QY 638 SSVADGR-IFQAGVSMGDCGCAQRNKPQVYTRLPFRDVI 677  
DB 237 --VTSKNIMWLIGDTSMGSCGCAKAYRPQYIGNVMTDVI 275

## RESULT 27

US-08-508-448C-25  
Sequence 25, Application US/08508448C

Patent No. 5804410

GENERAL INFORMATION:

APPLICANT: Kazuyoshi YAMAOKA et al.

TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING

FILE REFERENCE: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME

CURRENT APPLICATION NUMBER: 25

ADDRESS: Wenderoth, Lind & Ponack

STREET: 805 Fifteenth Street, N.W., #700

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/508.448C

FILING DATE: July 28, 1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-8850

TELEFAX:

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 418 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear  
US-08-508-448C-25

Query Match 13.5%; Score 511.5; DB 1; Length 418;  
Best Local Similarity 37.4%; Pred. No. 4.6e-29;  
Matches 108; Conservative 52; Mismatches 100; Indels 29; Gaps 10;

QY 409 SKNPEDCGKEDCDSDGDEK-----KDCDCGLRSFT-RQARVVGTDADGEGMPQVSL 460  
DB 145 NSGNLFINPSTETISLTDQAANWMLNECGAGPDLITLSBORLIGGTEABEGSWPQVSL 204  
QY 461 HALGQGHICGASLISPNMLVSAHACYIDDRGFYSDPTQMTAFGLHDG--QREAPVQVE 520  
DB 205 R-LNNAHCCGSLINNMWILTAHCFRSN-----SNPRMIATSGI-----STFPKLR-M 253  
QY 521 RLKRIISHPFNDPTDYDIALLEKPAEYSMWVPICLPDASHVFPAGKAIWVGW 580  
DB 254 RVNRIILHNNYSATENDIALVRLSNSVTFTDHSVCLPAATQNI PGSTAYVTGKGA 313  
QY 581 TOYGGTALILQGEIRVINQTTCCENLPOQ---ITPRMVCVGLSGVDSQCGDSGGL 636  
DB 314 QEYAGHTVPELRQGVRIISNDVCN--APHSYNGALISGMLCAGVPQGGVDACQDSGGL 371  
QY 637 LSSVADGR--IFQAGVSMGDCGCAQRNKPQVYTRLPFRDVIKENTGV 683  
DB 372 L--VQEDSRRLMPTVIGIVSMGDCGLPDKRGVYTRVATVLDWIRQDTGI 418

## RESULT 28

US-09-370-838-82  
Sequence 82, Application US/09370838

Patent No. 6444425

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Moses, Michael J.

APPLICANT: Mohamath, Roach

APPLICANT: Secrist, Heather

TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF

FILE REFERENCE: 210121.475C1

CURRENT APPLICATION NUMBER: US/09/370.838

CURRENT FILING DATE: 1999-08-09

EARLIER APPLICATION NUMBER: US 09/285,323

EARLIER FILING DATE: 1999-04-02

NUMBER OF SEQ ID NOS: 289

SOFTWARE: PaeSeq for Windows Version 3.0

SEQ ID NO 82

LENGTH: 418

TYPE: PRT

ORGANISM: Homo sapien

US-09-370-838-82

Query Match 13.5%; Score 511.5; DB 4; Length 418;  
Best Local Similarity 37.4%; Pred. No. 4.6e-29;  
Matches 108; Conservative 52; Mismatches 100; Indels 29; Gaps 10;

QY 409 SKNPEDCGKEDCDSDGDEK-----KDCDCGLRSFT-RQARVVGTDADGEGMPQVSL 460  
DB 145 NSGNLFINPSTETISLTDQAANWMLNECGAGPDLITLSBORLIGGTEABEGSWPQVSL 204  
QY 461 HALGQGHICGASLISPNMLVSAHACYIDDRGFYSDPTQMTAFGLHDG--QREAPVQVE 520  
DB 205 R-LNNAHCCGSLINNMWILTAHCFRSN-----SNPRMIATSGI-----STFPKLR-M 253  
QY 521 RLKRIISHPFNDPTDYDIALLEKPAEYSMWVPICLPDASHVFPAGKAIWVGW 580  
DB 254 RVNRIILHNNYSATENDIALVRLSNSVTFTDHSVCLPAATQNI PGSTAYVTGKGA 313  
QY 581 TOYGGTALILQGEIRVINQTTCCENLPOQ---ITPRMVCVGLSGVDSQCGDSGGL 636  
DB 314 QEYAGHTVPELRQGVRIISNDVCN--APHSYNGALISGMLCAGVPQGGVDACQDSGGL 371  
QY 637 LSSVADGR--IFQAGVSMGDCGCAQRNKPQVYTRLPFRDVIKENTGV 683



Db 372 L-VQEDSRRLMFIIVGIVSWGDCGLPDKPGVYTRVATYLDWIRQGTGI 418

## RESULT 29

US-09-370-838-83  
Sequence 83, Application US/09370838

Patent No. 6444425

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Lodes, Michael J.

APPLICANT: Mohamath, Radoch

APPLICANT: Secretier, Heather

TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF

TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE

FILE REFERENCE: 210121.475C1

CURRENT APPLICATION NUMBER: US/09/370.838

CURRENT FILING DATE: 1999-08-09

EARLIER APPLICATION NUMBER: US 09/285.323

NUMBER OF SEQ ID NOS: 289

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 83

LENGTH: 418

TYPE: PRT

ORGANISM: Homo sapien

US-09-370-838-83

Query Match 13.5%; Score 511.5; DB 4; Length 418;  
Best Local Similarity 37.4%; Pred. No. 4.6e-29;  
Matches 108; Conservative 52; Mismatches 100; Indels 29; Gaps 10;

Qy 409 SKNPECDGKEDSGDSDE-----KDCDGLRSFT-RQARVVGCTDADGEMPMQVSL 460  
Db 145 NSGNLEINPSTETISLTDOAAANWLINECAGPDLITLSEQRILGTEAEBSMPQVSL 204  
Qy 461 HALGQGHICGASLISPMVLVSAHCYIDRGFRYSDEPTQWTAFLGLHDQSORAPGVQER 520  
Db 205 R-LNNAHCGSSLINNMWILTAHCFRSN-----SNPRDWIATSGI-----STFPKLR-M 253  
Qy 521 RLRRIISHPPFNDFTFYDIALLEKPAEYSSMWPRICLPDASHVFPAGKAIWYTGWGH 580  
Db 254 RVRNIIHNNYKSATHEINDIALVRLNLSVTFKDIHVCIPAAATONIPPGSTAYVTGWA 313  
Qy 581 TOYGTGALLIOLKEIRVINTTCENLPPQ-----ITPRMVCVGLSGVDSGCGDSGCP 636  
Db 314 QEYAGHTVPELRQGVRIISNDVCN--APHSYNGALISGMLCAGVPQGVADACGDSGCP 371  
Qy 637 LSSVEADGR--IFQAGVSWGDCGCAQRNKPQVYTRPLFRDWIKENTGV 683  
Db 372 L-VQEDSRRLMFIIVGIVSWGDCGLPDKPGVYTRVATYLDWIRQGTGI 418

## RESULT 30

US-09-854-133-82  
Sequence 82, Application US/09854133

Patent No. 6759508

GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.

APPLICANT: Mohamath, Radoch

APPLICANT: Henderson, Robert A.

APPLICANT: Benson, Darin R.

APPLICANT: Secretier, Heather

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.475C10

CURRENT APPLICATION NUMBER: US/09/854.133

NUMBER OF SEQ ID NOS: 735

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 82

LENGTH: 418

TYPE: PRT

ORGANISM: Homo sapien

Query Match 13.5%; Score 511.5; DB 4; Length 418;  
Best Local Similarity 37.4%; Pred. No. 4.6e-29;  
Matches 108; Conservative 52; Mismatches 100; Indels 29; Gaps 10;

Qy 409 SKNPECDGKEDSGDSDE-----KDCDGLRSFT-RQARVVGCTDADGEMPMQVSL 460  
Db 145 NSGNLEINPSTETISLTDOAAANWLINECAGPDLITLSEQRILGTEAEBSMPQVSL 204  
Qy 461 HALGQGHICGASLISPMVLVSAHCYIDRGFRYSDEPTQWTAFLGLHDQSORAPGVQER 520  
Db 205 R-LNNAHCGSSLINNMWILTAHCFRSN-----SNPRDWIATSGI-----STFPKLR-M 253  
Qy 521 RLRRIISHPPFNDFTFYDIALLEKPAEYSSMWPRICLPDASHVFPAGKAIWYTGWGH 580  
Db 254 RVRNIIHNNYKSATHEINDIALVRLNLSVTFKDIHVCIPAAATONIPPGSTAYVTGWA 313  
Qy 581 TOYGTGALLIOLKEIRVINTTCENLPPQ-----ITPRMVCVGLSGVDSGCGDSGCP 636  
Db 314 QEYAGHTVPELRQGVRIISNDVCN--APHSYNGALISGMLCAGVPQGVADACGDSGCP 371  
Qy 637 LSSVEADGR--IFQAGVSWGDCGCAQRNKPQVYTRPLFRDWIKENTGV 683  
Db 372 L-VQEDSRRLMFIIVGIVSWGDCGLPDKPGVYTRVATYLDWIRQGTGI 418

Search completed: November 29, 2004, 08:34:26  
Job time: 30.7568 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: November 29, 2004, 08:14:53 ; Search time 98.1424 Seconds

(without alignments)  
2496.496 Million cell updates/sec

Title: US-09-936-333-5

Perfect score: 3789

Sequence: 1 MAERVYMLPPRRASLSKSFV.....PGVYTRLP.LPRDKENTGV 683

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing-First-100 summaries

Database :

A\_Geneseq\_23Sep04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3760	99.2	683	3 AAB19551	Aab19551 Human mat
2	3760	99.2	762	3 AAY90284	Aay90284 Human pep
3	3760	99.2	855	3 AAB19552	Aab19552 Human mat
4	3760	99.2	855	4 AAB35465	Aab35465 Human mem
5	3760	99.2	855	5 ADI16817	Adi16817 Human NOV
6	3760	99.2	855	5 ADI16883	Adi16883 Human NOV
7	3760	99.2	855	5 ADI16876	Adi16876 Human NOV
8	3760	99.2	855	7 ADN39867	Adn39867 Cancer/an
9	3760	99.2	855	8 ADN04754	Adn04754 Anticpctri
10	3756	99.1	855	5 ADI16884	Adi16884 Human NOV
11	3756	99.1	855	5 ADI16818	Adi16818 Human NOV
12	3754	99.1	757	5 ADI16508	Adi16508 Human NOV
13	3754	99.1	855	2 AAY06671	Aay06671 Tumour an
14	3754	99.1	855	4 AAB98500	Aab98500 Human TRAD
15	3754	99.1	855	4 AAE06930	Aae06930 Human mem
16	3754	99.1	855	5 AAO22929	Aao22929 Type II t
17	3754	99.1	855	5 ADI16816	Adi16816 Human NOV
18	3754	99.1	855	5 ADI16882	Adi16882 Human NOV
19	3754	99.1	855	5 ADI16875	Adi16875 Human NOV
20	3754	99.1	855	6 ABB56619	Abb56619 Human mem
21	3754	99.1	855	6 AAO30146	Aao30146 Human mem
22	3754	99.1	855	6 AAE29820	Aae29820 Human mem
23	3754	99.1	855	6 AAE29791	Aae29791 Human mem
24	3754	99.1	855	6 ABB72376	Abb72376 Transmemb
25	3754	99.1	855	7 ADB97551	Adb97551 Human MTS

26	3754	99.1	855	7 ADI10371	Adi10371 Human cel
27	3754	99.1	855	8 ADG65326	Adg65326 Human MTS
28	3754	99.1	855	8 ADI28861	Adi28861 Human mat
29	3754	99.1	855	8 ADJ46895	Adj46895 Human tra
30	3745	98.8	851	4 AAM25628	Aam25628 Human pro
31	3745	98.8	851	4 AAB11428	Aab11428 Human mem
32	3726	98.3	853	8 ADO55145	Ado55145 Protein #
33	3719	98.2	757	8 ADN45162	Adn45162 Human nov
34	3549	93.7	922	4 ABB31442	Abb31442 Novel hum
35	3259.5	86.0	782	5 ABB96427	Abb96427 Human ova
36	3209	84.7	855	5 AAE23083	Aae23083 Epithin p
37	3209	84.7	855	5 ADI16819	Adi16819 Murine NO
38	3209	84.7	855	5 ADI16877	Adi16877 Murine NO
39	3197	84.4	855	5 ADI16820	Adi16820 Rat NOVX
40	3197	84.4	855	5 ADI16881	Adi16881 Rat NOVX
41	3197	84.4	855	5 ADI16878	Adi16878 Rat NOVX
42	3123	82.4	902	4 AAB98507	Aab98507 Murine ep
43	3123	82.4	902	5 AAB080517	Aab080517 Mouse epi
44	3123	82.4	902	5 AAU77549	Aau77549 Murine ty
45	2797	73.8	620	3 AAB43748	Aab43748 Human can
46	2254.5	59.5	845	5 AAB16879	Abi16879 African c
47	1319	34.8	241	4 AAE06936	Aae06936 Human mem
48	1319	34.8	241	5 AAO22930	Aao22930 Type II t
49	1319	34.8	241	5 AAE22837	Aae22837 Human mat
50	1319	34.8	241	6 ABB56620	Abb56620 Human MTS
51	1319	34.8	241	6 AAO30147	Aao30147 Human mem
52	1319	34.8	241	6 AAE29821	Aae29821 Human mem
53	1319	34.8	241	6 AAE29792	Aae29792 Human mem
54	1319	34.8	241	6 ABB72377	Abb72377 Transmemb
55	1319	34.8	241	7 ADA25176	Ada25176 Human MTS
56	1319	34.8	241	7 ADB97553	Adb97553 Human MTS
57	1319	34.8	241	8 ADI28862	Adi28862 Human mat
58	1315	34.7	241	2 AAW22987	Aaw22987 Human ser
59	1025	27.2	850	5 AAU82750	Aau82750 Amino aci
60	1026	27.1	693	7 AAE38320	Aae38320 Human mem
61	1020	26.9	648	7 AAE38322	Aae38322 Human mem
62	1019.5	26.9	706	7 AAE38321	Aae38321 Human mem
63	1019.5	26.9	843	5 AAU77552	Aau77552 Human memb
64	1019.5	26.9	843	5 AAE38319	Aae38319 Human memb
65	1015.5	26.8	572	8 ADG31207	Adg31207 Novel mou
66	1014	26.8	799	5 ADI16874	Adi16874 Murine NO
67	1014	26.8	799	5 ADI16880	Adi16880 Murine NO
68	1007	26.6	795	5 ABB43952	Abb43952 Human PRO
69	1005	26.5	802	2 AAY41710	Aay41710 Human PRO
70	1005	26.5	802	3 AAB44266	Aab44266 Human PRO
71	1005	26.5	802	3 AAB24052	Aab24052 Human PRO
72	1005	26.5	802	5 AAB82755	Aab82755 Amino aci
73	1005	26.5	802	6 ABO25212	Abo25212 Novel hum
74	1005	26.5	802	6 ABO72218	Abo72218 Novel hum
75	1005	26.5	802	6 ABO84898	Abo84898 Human sec
76	1005	26.5	802	6 ABO61096	Abo61096 Human sec
77	1005	26.5	802	6 ABO80365	Abo80365 Human sec
78	1005	26.5	802	6 ADA24708	Ada24708 Novel hum
79	1005	26.5	802	6 ABO19667	Abo19667 Novel hum
80	1005	26.5	802	6 ADA12369	Ada12369 Human sec
81	1005	26.5	802	6 ABO19558	Abo19558 Novel hum
82	1005	26.5	802	7 ADB73675	Adb73675 Human PRO
83	1005	26.5	802	7 ADB76391	Adb76391 Human PRO
84	1005	26.5	802	7 ADC43817	Adc43817 Human sec
85	1005	26.5	802	7 ADC61577	Adc61577 Human sec
86	1005	26.5	802	7 ADC63541	Adc63541 Human sec
87	1005	26.5	802	7 ADC65417	Adc65417 Human sec
88	1005	26.5	802	7 ADC66641	Adc66641 Human sec
89	1005	26.5	802	7 ADC68765	Adc68765 Human sec
90	1005	26.5	802	7 ADC68285	Adc68285 Human sec
91	1005	26.5	802	7 ADC67890	Adc67890 Human sec
92	1005	26.5	802	7 ADC41210	Adc41210 Human sec
93	1005	26.5	802	7 ADC67265	Adc67265 Human sec
94	1005	26.5	802	7 ADC62201	Adc62201 Human sec
95	1005	26.5	802	7 ADC41834	Adc41834 Human sec
96	1005	26.5	802	7 ADB849203	Adb849203 Human sec
97	1005	26.5	802	7 ADB835257	Adb835257 Human sec
98	1005	26.5	802	7 ADB16371	Adb16371 Human sec
99	1005	26.5	802	7 ADD72986	Add72986 Human sec



QY 481 SAACHYIDDRGFRYSDDPTQMTAFILGLHDQSORSAPGVQERLKRILISHPFENDFTFYDI 540  
DB 481 SAACHYIDDRGFRYSDDPTQMTAFILGLHDQSORSAPGVQERLKRILISHPFENDFTFYDI 540  
QY 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGNGHTQYGGTGLILQKEIRVIN 600  
DB 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGNGHTQYGGTGLILQKEIRVIN 600  
QY 601 QTTCENTLLPQOITPRMNCVGFSLSGVDSCCGDSGGPLSVYADGRIFQAGVSWGDCAQ 660  
DB 601 QTTCENTLLPQOITPRMNCVGFSLSGVDSCCGDSGGPLSVYADGRIFQAGVSWGDCAQ 660  
QY 661 RNKPGVYTRLPFLRDMIKENTGV 683  
DB 661 RNKPGVYTRLPFLRDMIKENTGV 683

RESULT 2  
AAV90284  
ID AAV90284 standard; protein; 762 AA.  
AC AAV90284;  
DT 24-OCT-2000 (first entry)  
XX Human peptidase, HPEP-1 protein sequence.  
DE Human peptidase, HPEP-1 protein sequence.  
XX Human; peptidase; cell proliferative disorder; arteriosclerosis;  
KM prolatitis; myelofibrosis; cancer; autoimmune disorder; Crohn's disease;  
KM inflammatory disorder; AIDS; anaemia; allergy; asthma; arteriosclerosis;  
KM Grave's disease; multiple sclerosis; scleroderma; infection; diabetes;  
KM metabolic disorder; Addison's disease; cystic fibrosis; diagnosis;  
KM glycogen storage disease; obesity; therapy; HPEP-1.  
XX Homo sapiens.  
OS Homo sapiens.  
PN WO200042201-A2.  
XX 20-JUL-2000.  
PF 11-JAN-2000; 2000WO-US000641.  
PR 11-JAN-1999; 99US-0172247P.  
PR 03-MAY-1999; 99US-0132253P.  
PR 27-MAY-1999; 99US-013653P.  
XX (INCY-) INCYTE PHARM INC.  
PA Bandman O, Hillman JL, Tang YT, Azimzai Y, Baughn MR, Lal P;  
PI Yue H, Lu DM;  
XX WPI; 2000-482832/42.  
DR N-PSDB; AAA37657.  
XX An isolated polypeptide for diagnosis, prevention and treatment of cell  
PT proliferative, autoimmune/ inflammatory and metabolic disorders comprises  
PT a sequence encoding a human peptidase.  
PS Claim 2; Page 91-93; 131pp; English.

XX SQ Sequence 762 AA:  
Query Match 99.2%; Score 3760; DB 3; Length 762;  
Best Local Similarity 99.4%; Pred. No. 4.1e-241;  
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MAERVYMLPPRARSLKSFVTVTSVAEPPTDSTKVTORTDSCSGFLHARGVELMRFTTPG 60  
DB 80 MAERVYMLPPRARSLKSFVTVTSVAEPPTDSTKVTORTDSCSGFLHARGVELMRFTTPG 139  
QY 61 FPDSPYAHARCCWALRGDADSVLSLTFRSFDLASCDERSDLYVYNTLSPEPHALVQ 120  
DB 140 FPDSPYAHARCCWALRGDADSVLSLTFRSFDLASCDERSDLYVYNTLSPEPHALVQ 139  
QY 121 LCGTYPSPSYNLTHRSSQNVLLITLITTERRRHGFEXTFQOLRMSCGRLKAQCTFN 180  
DB 200 LCGTYPSPSYNLTHRSSQNVLLITLITTERRRHGFEXTFQOLRMSCGRLKAQCTFN 259  
QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVRFKPFYILEPRAACGCPDYVEINGEKYCGE 240  
DB 260 SPYYPGHYPNIDCTWNIIEVNNQHVRFKPFYILEPRAACGCPDYVEINGEKYCGE 319  
QY 241 RSQFVYTSNSNKTIVRFHSDQSYTDGFLAEVLSYSSDPCPGQFTCRGCIKELRCD 300  
DB 320 RSQFVYTSNSNKTIVRFHSDQSYTDGFLAEVLSYSSDPCPGQFTCRGCIKELRCD 379  
QY 301 GMADCTHSDBLNCSCDAGHQFTCKKFKCYPLWVCDNVNDCGDNSEDGSCCPAOTFR 360  
DB 380 GMADCTHSDBLNCSCDAGHQFTCKKFKCYPLWVCDNVNDCGDNSEDGSCCPAOTFR 439  
QY 361 SNGKCLSKSQOQNGKXDCGSDSDEASCPRYNVYCTGKTRCLNGCLSKGNPECDGKED 420  
DB 440 SNGKCLSKSQOQNGKXDCGSDSDEASCPRYNVYCTGKTRCLNGCLSKGNPECDGKED 499  
QY 421 CSDGSEKDCDCGLRSTFQARVVGTDADGEMPMQVSLHALQGGHICASLISPMVLV 480  
DB 500 CSDGSEKDCDCGLRSTFQARVVGTDADGEMPMQVSLHALQGGHICASLISPMVLV 559  
QY 481 SAACHYIDDRGFRYSDDPTQMTAFILGLHDQSORSAPGVQERLKRILISHPFENDFTFYDI 540  
DB 560 SAACHYIDDRGFRYSDDPTQMTAFILGLHDQSORSAPGVQERLKRILISHPFENDFTFYDI 619  
QY 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGNGHTQYGGTGLILQKEIRVIN 600  
DB 620 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGNGHTQYGGTGLILQKEIRVIN 679  
QY 601 QTTCENTLLPQOITPRMNCVGFSLSGVDSCCGDSGGPLSVYADGRIFQAGVSWGDCAQ 660  
DB 680 QTTCENTLLPQOITPRMNCVGFSLSGVDSCCGDSGGPLSVYADGRIFQAGVSWGDCAQ 739  
QY 661 RNKPGVYTRLPFLRDMIKENTGV 683  
DB 740 RNKPGVYTRLPFLRDMIKENTGV 762

RESULT 3  
AAB19552  
ID AAB19552 standard; protein; 855 AA.  
AC AAB19552;  
DT 22-JAN-2001 (first entry)  
XX Human matrixase.  
DE Human matrixase.  
XX Matrixase; serine protease; human; breast cancer; pre-malignancy;  
KM actinic keratosis; leukoplakia; Barrett's epithelium;  
KM columnar metaplasia; ulcerative colitis; Bowenoid papulosis;  
KM adenomatous colorectal polyp; Ovarian erythroplasia;  
KM vulvar intraepithelial neoplasia; tumour; metastasis; therapy.  
OS Homo sapiens.

XX	Key	Location/Qualifiers
FT	Region	214..327
FT	Peptide	/note= "complement subcomponent 1r and 1s (c1r/s) region"
FT	Modified-site	249..251
FT	Modified-site	/note= "RCD motif"
FT	Region	302
FT	Region	/note= "N-glycosylated"
FT	Region	340..440
FT	Region	/note= "complement subcomponent 1r and 1s (c1r/s) region"
FT	Region	452..486
FT	Modified-site	/note= "IDL-receptor type region"
FT	Modified-site	485
FT	Region	/note= "N-glycosylated"
FT	Region	487..523
FT	Region	/note= "IDL-receptor type region"
FT	Region	524..559
FT	Region	/note= "IDL-receptor type region"
FT	Region	566..602
FT	Domain	/note= "IDL-receptor type region"
FT	Cleavage-site	603..855
FT	Cleavage-site	/label= Catalytic_domain
FT	Cleavage-site	614..618
FT	Cleavage-site	/note= "conserved proteolytic activation site"
PN	WO200053232-A1.	
XX	14-SEP-2000.	
PD	10-MAR-2000; 2000WO-US006111.	
XX	12-MAR-1999; 99US-0124006P.	
PF	(GEOU ) UNIV GEORGETOWN.	
XX	Dickson RB, Lin C, Johnson M, Wang S, Enyed I;	
XX	WPI: 2000-594268/56.	
XX	N-PDSB; AAA86493.	
DR	Treating malignancies, premalignant and pathologic conditions in a	
XX	subject, comprises administering matrixase modulating agent.	
PT	Claim 14; Fig 16; 116pp; English.	
XX	The present sequence is that of the full-length form of human matrixase	
CC	a trypsin-like protease. This has an additional 172 amino acids compared	
CC	with the truncated form of matrixase given in AAB19551. Either form can	
CC	be produced in transformed or transfected cells using a claimed method.	
CC	The zymogen (inactive) form of matrixase is a single-chain protein. The	
CC	active 2-chain form strongly interacts with fragments of a Kunitz-type	
CC	serine protease inhibitor (hepatocyte growth factor activator inhibitor,	
CC	HAI-1) to form SDS-stable complexes. In breast cancer cells, matrixase	
CC	is present mainly as the uncomplexed form. Only the complexed matrixase	
CC	is detected in human milk. The invention is directed to a method of	
CC	detecting a malignancy or a pre-malignant lesion in breast or other	
CC	tissue by detecting the presence of single- or 2-chain forms of	
CC	matrixase in the tissue. The object is to inhibit tumour onset, tumour	
CC	growth and metastasis. Malignancies and pre-malignant conditions	
CC	characterised by expression of the zymogen or activated form of	
CC	matrixase are treated by administering an inhibitor of matrixase,	
CC	especially a Bowman-Birk inhibitor. The pre-malignant condition is	
CC	atypical ductal hyperplasia of the breast, actinic keratosis,	
CC	leukoplakia, Barrett's epithelium of the esophagus, ulcerative colitis,	
CC	adenomatous colorectal polyps, erythroplasia of the Queyrat, Bowen's	
CC	disease, Bowenoid papulosis, vulvar intraepithelial neoplasia or	
CC	dyplastic changes to the cervix. The invention also provides methods for	
CC	in vivo or in vitro diagnosis of malignancy or pre-malignant lesion, and	
CC	methods of identifying matrixase modulators, including activators and	
CC	inhibitors	
XX	Sequence 855 AA;	
50		

Query Match	Similarity	99.2%;	Score 3760;	DB 3;	Length 855;			
Best Local	Similarity	99.4%;	Pred. No. 4.7e-241;					
Matches	Conservative	0;	Mismatches	4;	Indels	0;	Gaps	0;
QY	1	MAEERVYMLPPRARSLKSFVVTSVVAPEPTDSKVQRTDNCSCFGLHARGVELMKEFTTTPG	60					
DB	173	MAEERVYMLPPRARSLKSFVVTSVVAPEPTDSKVQRTDNCSCFGLHARGVELMKEFTTTPG	232					
QY	61	FPDSPPYPAHACQWALRGDADSVSLTFPRSGDLASCPERGSDLTVYVNTLSPMEPHALVO	120					
DB	233	FPDSPPYPAHACQWALRGDADSVSLTFPRSGDLASCPERGSDLTVYVNTLSPMEPHALVO	292					
QY	121	LCGYTPESYNLTFHSSQNVLLITLITINTERHPGFEATFQLPRMSSCGRLRKAQGTFN	180					
DB	293	LCGYTPESYNLTFHSSQNVLLITLITINTERHPGFEATFQLPRMSSCGRLRKAQGTFN	352					
QY	181	SPYYPGHYPNIDICTWANIENVNNQHKYRFEFFLLERPRACGTCPCPKDYVINGEKYGE	240					
DB	353	SPYYPGHYPNIDICTWANIENVNNQHKYRFEFFLLERPRACGTCPCPKDYVINGEKYGE	412					
QY	241	RSQGVYVNSNSKITYRFPHSDSYNDTGFLAFLSYDSSDPGQFTCRGCRIRKELRCD	300					
DB	413	RSQGVYVNSNSKITYRFPHSDSYNDTGFLAFLSYDSSDPGQFTCRGCRIRKELRCD	472					
QY	301	GMADCTHSDSLNCSCDAGHQFTCKNFKCKPLFWVCDSVNDGDNSEDQSCSPAQTFRC	360					
DB	473	GMADCTHSDSLNCSCDAGHQFTCKNFKCKPLFWVCDSVNDGDNSEDQSCSPAQTFRC	532					
QY	361	SNGCLSLKSQOCNKDCCGSDSDASCPKXNVVYCTGHTYRCLNGLICLSKGNPCDGKED	420					
DB	533	SNGCLSLKSQOCNKDCCGSDSDASCPKXNVVYCTGHTYRCLNGLICLSKGNPCDGKED	592					
QY	421	CSDSDSEKDCDCGLRSFTROARVVGTTDADBGEMPMQVSLHALQGHICGASLISPNMLV	480					
DB	593	CSDSDSEKDCDCGLRSFTROARVVGTTDADBGEMPMQVSLHALQGHICGASLISPNMLV	652					
QY	481	SAACCTYIDDRFRISDPQTQWTAFLGLHDQSGRSAPGVQERLKLIISSHPFNDFPVDI	540					
DB	653	SAACCTYIDDRFRISDPQTQWTAFLGLHDQSGRSAPGVQERLKLIISSHPFNDFPVDI	712					
QY	541	ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGWHQYGGTGALLIQKGBIRVIN	600					
DB	713	ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGWHQYGGTGALLIQKGBIRVIN	772					
QY	601	QTTCECNLLPQOITPRMNCVGLSGGVSDSCGDSGGPLSSVEADGRIFQAGVYVSWGDCGQ	660					
DB	773	QTTCECNLLPQOITPRMNCVGLSGGVSDSCGDSGGPLSSVEADGRIFQAGVYVSWGDCGQ	832					
QY	661	RNKGQVYTRLPFLFRDWIKENTGV	683					
DB	833	RNKGQVYTRLPFLFRDWIKENTGV	855					
RESULT 4								
AAB35465								
ID AAB35465 standard; protein, 855 AA.								
XX	AC	AAB35465;						
XX	DT	06-JUN-2001 (first entry)						
XX	DE	Human membrane-type serine protease MT-SPL.						
XX	KW	Human; membrane-type serine protease; MT-SPL; cancer.						
XX	OS	Homo sapiens.						
XX	PN	MO200123524-A2.						
XX	PD	05-APR-2001.						
XX	PF	02-OCT-2000; 2000MO-US027250.						

PR 30-SEP-1999; 99US-00410362.  
XX (REGC ) UNIV CALIFORNIA.  
XX Craik CS, Takeuchi T, Shuman M;  
XX WPI; 2001-245002/25.  
DR N-PSDB; AAF28099.  
XX  
XX New nucleic acid encoding a membrane type serine protease, useful for the  
PT diagnosis, prognosis and treatment of cancer, particularly metastatic  
PT cancers.  
XX  
PS Claim 7; Fig 1; 102pp; English.  
XX  
XX The present invention provides the protein and coding sequences for the  
CC novel human membrane-type serine protease MT-Sp1. Increased expression of  
CC this protein is associated with cancer, and so the sequences can be used  
CC in cancer diagnosis and the identification of treatments. The present  
CC sequence is the MT-Sp1 protein  
XX  
XX Sequence 855 AA;  
SQ  
Query Match 99.2%; Score 3760; DB 4; Length 855;  
Best Local Similarity 99.4%; Pred. No. 4,76-241;  
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MAEERVVMLPPRARSLKSPVTVSVAPPTDSKTVQRTQDNCSFGLHARGVELMRFTTPG 60  
Db 173 MAEERVVMLPPRARSLKSPVTVSVAPPTDSKTVQRTQDNCSFGLHARGVELMRFTTPG 232  
QY 61 PPDSPPYAHARCOVALGDDADSVSLTPRSFDLASCDERGDLYTVYVNTLSPMEPHALVQ 120  
Db 233 PPDSPPYAHARCOVALGDDADSVSLTPRSFDLASCDERGDLYTVYVNTLSPMEPHALVQ 292  
QY 121 LCGTYPSPSYNLTFPSSQNVLLITLITNTERRHPGFEATFPLPMNSCGGLRKAQGTFN 180  
Db 293 LCGTYPSPSYNLTFPSSQNVLLITLITNTERRHPGFEATFPLPMNSCGGLRKAQGTFN 352  
QY 181 SPYYPGHYPNIDCTWNIIEVBNQHVKKRFPYLLBPRACGTCPKDYVEINEKTCGE 240  
Db 353 SPYYPGHYPNIDCTWNIIEVBNQHVKKRFPYLLBPRACGTCPKDYVEINEKTCGE 412  
QY 241 RSQVTVTSNSNKTIVRRHSDSDSYDTGTGLAFLSYDSSDPFGQFTCTGTCIKELACD 300  
Db 413 RSQVTVTSNSNKTIVRRHSDSDSYDTGTGLAFLSYDSSDPFGQFTCTGTCIKELACD 472  
QY 301 GMADCTHSDSLNCSCDAGHOFCKNKPCKPLFWVCDVNDGNSDDEGSCCPAQTPRC 360  
Db 473 GMADCTHSDSLNCSCDAGHOFCKNKPCKPLFWVCDVNDGNSDDEGSCCPAQTPRC 532  
QY 361 SNGKCLSKSQCNKGDCGSDDEASCPKVVVYCTKTKTYRCLNGLCLSKGNPECDKED 420  
Db 533 SNGKCLSKSQCNKGDCGSDDEASCPKVVVYCTKTKTYRCLNGLCLSKGNPECDKED 592  
QY 421 CSDGSDDEDCGCGLRSTFTROARVVGCTDADGEMPMQVSLHALQGHICGASLISPMNLV 480  
Db 593 CSDGSDDEDCGCGLRSTFTROARVVGCTDADGEMPMQVSLHALQGHICGASLISPMNLV 652  
QY 481 SAACHYIDDRGFRYSDFQTWTAFLGLHDQSORSAAGVERLKRILISHPFNDFTFYDI 540  
Db 653 SAACHYIDDRGFRYSDFQTWTAFLGLHDQSORSAAGVERLKRILISHPFNDFTFYDI 712  
QY 541 ALLELEKRAEYSSWVRPICLPDASHVPAGRAIWTGNGHTQYGGTGLILQKEIRVIN 600  
Db 713 ALLELEKRAEYSSWVRPICLPDASHVPAGRAIWTGNGHTQYGGTGLILQKEIRVIN 772  
QY 601 QTTCEMLLPQITTRAMCVGLSGVDSCQDSSGGLSSVADGRIFPAGVIVSMGDCAQ 660  
Db 773 QTTCEMLLPQITTRAMCVGLSGVDSCQDSSGGLSSVADGRIFPAGVIVSMGDCAQ 832  
QY 661 RNKPGVYTRLPFRDWIKENTGV 683

Db 833 RNKPGVYTRLPFRDWIKENTGV 855  
RESULT 5  
AD116817  
ID AD116817 standard; protein; 855 AA.  
XX  
XX AD116817;  
AC  
XX 15-APR-2004 (first entry)  
DT  
XX Human NOXV protein homologue SeqID 353.  
DE  
XX  
XX human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;  
KW inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; achmia;  
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; str.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200268649-A2.  
FN  
XX  
XX 06-SEP-2002.  
PD  
XX 31-JAN-2002; 2002MO-US002785.  
PF  
XX  
XX 31-JAN-2001; 2001US-0265395P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266757P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 07-FEB-2001; 2001US-0267057P.  
PR 08-FEB-2001; 2001US-0267459P.  
PR 09-FEB-2001; 2001US-0267823P.  
PR 15-FEB-2001; 2001US-0268974P.  
PR 26-FEB-2001; 2001US-0271664P.  
PR 27-FEB-2001; 2001US-0271839P.  
PR 27-FEB-2001; 2001US-0271855P.  
PR 02-MAR-2001; 2001US-0272788P.  
PR 02-MAR-2001; 2001US-0273406P.  
PR 14-MAR-2001; 2001US-0275925P.  
PR 14-MAR-2001; 2001US-0275947P.  
PR 14-MAR-2001; 2001US-0275950P.  
PR 14-MAR-2001; 2001US-0275959P.  
PR 15-MAR-2001; 2001US-0276448P.  
PR 15-MAR-2001; 2001US-0276450P.  
PR 16-MAR-2001; 2001US-0276397P.  
PR 16-MAR-2001; 2001US-0276768P.  
PR 20-MAR-2001; 2001US-0278652P.  
PR 26-MAR-2001; 2001US-0278775P.  
PR 26-MAR-2001; 2001US-0278778P.  
PR 29-MAR-2001; 2001US-0279882P.  
PR 29-MAR-2001; 2001US-0279884P.  
PR 30-MAR-2001; 2001US-0280147P.  
PR 11-APR-2001; 2001US-0282932P.  
PR 11-APR-2001; 2001US-0283083P.  
PR 20-APR-2001; 2001US-0285133P.  
PR 23-APR-2001; 2001US-0285749P.  
PR 03-MAY-2001; 2001US-0286137P.  
PR 03-MAY-2001; 2001US-0288504P.  
PR 29-MAY-2001; 2001US-0294047P.  
PR 30-MAY-2001; 2001US-0294734P.  
PR 08-JUN-2001; 2001US-0296964P.  
PR 18-JUN-2001; 2001US-0298959P.  
PR 19-JUN-2001; 2001US-0299334P.  
PR 13-AUG-2001; 2001US-0312020P.  
PR 16-AUG-2001; 2001US-0312889P.  
PR 16-AUG-2001; 2001US-0312908P.  
PR 21-AUG-2001; 2001US-0313390P.  
PR 28-AUG-2001; 2001US-0315470P.



PR 31-AUG-2001; 2001US-0316447P.  
PR 07-SEP-2001; 2001US-0318115P.  
PR 07-SEP-2001; 2001US-0318118P.  
PR 12-SEP-2001; 2001US-0318740P.  
PR 19-SEP-2001; 2001US-0323379P.  
PR 18-OCT-2001; 2001US-0320245P.  
PR 18-OCT-2001; 2001US-0330308P.  
PR 14-NOV-2001; 2001US-0332701P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Tchernev VT, Szytek KA, Zernhusen BD, Patturajan M, Shinkets RA;  
PI Li L, Ganggoli EA, Radigaru M, Anderson DW, Rastelli L, Miller CE;  
PI Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CE;  
PI Futrak K, Grosse WM, Alebrock JP, Lepley DM, Rieger DK, Burgess CE;  
XX  
XX WPI; 2002-706998/76.  
XX  
XX New NOXV polypeptides and nucleic acids, useful for preventing or  
PT treating NOXV-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX  
XX Disclosure; SEQ ID NO 353; 1498bp; English.  
XX  
XX This invention relates to a novel nucleic acids, and encoded polypeptides  
CC thereof, which have properties related to the stimulation of biochemical  
CC or physiological responses in a cell, tissue, organ or organism.  
CC Specifically, it refers to the use of biologically active fragments for  
CC diagnostic and prognostic assays and furthermore in the treatment of  
CC diverse pathological conditions. The present invention describes novel  
CC human and murine NOXV proteins, as well as methods to modulate their  
CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
CC The NOXV polypeptides, polynucleotides and antibodies are useful in  
CC treating or preventing NOXV-associated disorders, e.g. cardiomyopathy,  
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
CC treating or preventing diseases such as inflammation, autoimmune  
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,  
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
CC and epilepsy. Accordingly, these molecules have many activities including  
CC cytostatic, cariant, antiinflammatory, immunosuppressive, anti-allergic,  
CC hemostatic, anti-HIV, antidiabetic, antiatherosclerotic, anorectic,  
CC antiaesthetic, nephrotoxic, antiaerobic, hepatocytic,  
CC neuroprotective, nootropic, antidiabetic, vitruce, antiparasitic,  
CC relaxant and anticonvulsant. In addition, they are useful in screening  
CC assays to identify small molecules that modulate or inhibit, for example,  
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
CC used as in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. This polypeptide is a homologue of a human NOXV protein  
CC of the invention.  
XX  
XX Sequence 855 AA:  
SQ  
Query Match 99.2%; Score 3760; DB 5; Length 855;  
Best Local Similarity 99.4%; Pred. No. 4,7e-241;  
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 241 RSQFVVTNSNKTITVAFHSDQSYTDTGFLAEYLISYSDSPCPOQFTCRTRCIRKELRCD 300  
DB 413 RSQFVVTNSNKTITVAFHSDQSYTDTGFLAEYLISYSDSPCPOQFTCRTRCIRKELRCD 472  
QY 301 GMAADCTDHSDELNCSQDAGHOFCKKFKCPLFWVCDSDVNDCCGNSDEQSCSPAQFRFC 360  
DB 473 GMAADCTDHSDELNCSQDAGHOFCKKFKCPLFWVCDSDVNDCCGNSDEQSCSPAQFRFC 532  
QY 361 SNKCKLSKSGQCGKDDCGSDGDEASCPKNVVTCTKHTRCNLGCLSGNPECDEKED 420  
DB 533 SNKCKLSKSGQCGKDDCGSDGDEASCPKNVVTCTKHTRCNLGCLSGNPECDEKED 592  
QY 421 CSDGSEKDCDCGIRSFTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480  
DB 593 CSDGSEKDCDCGIRSFTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652  
QY 481 SAACHYIDRGRFYSPTQMTAFGLHDSQSBAPQVQERLKRITISHPFNDFTPYDI 540  
DB 653 SAACHYIDRGRFYSPTQMTAFGLHDSQSBAPQVQERLKRITISHPFNDFTPYDI 712  
QY 541 ALLELEKPAEYSMBRPICLPDASHVFPAGKAIWTVGMGHTQYGGTALILQGEIRVIN 600  
DB 713 ALLELEKPAEYSMBRPICLPDASHVFPAGKAIWTVGMGHTQYGGTALILQGEIRVIN 772  
QY 601 QTTCEMLPQOITPRMNCVGLSGVDSQCGDSGGLSYEADGRIFQAGVSMGDCAQ 660  
DB 773 QTTCEMLPQOITPRMNCVGLSGVDSQCGDSGGLSYEADGRIFQAGVSMGDCAQ 832  
QY 661 RNRPGVYTRPLPRDINKENTGV 683  
DB 833 RNRPGVYTRPLPRDINKENTGV 855  
RESULT 6  
ID AD116883 standard; protein; 855 AA.  
XX  
XX AD116883;  
XX  
DT 15-Apr-2004 (first entry)  
XX  
XX Human NOXV protein homologue SegID 419.  
XX  
XX human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;  
KW inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; stc.  
XX  
XX Homo sapiens.  
XX  
XX W0200268649-A2.  
XX  
PD 06-SEP-2002.  
XX  
PF 31-JAN-2002; 2002WO-US002785.  
XX  
PR 31-JAN-2001; 2001US-0265195P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266767P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 07-FEB-2001; 2001US-0267057P.  
PR 08-FEB-2001; 2001US-0267459P.  
PR 09-FEB-2001; 2001US-0267823P.  
PR 15-FEB-2001; 2001US-0268974P.  
PR 26-FEB-2001; 2001US-0271664P.  
PR 27-FEB-2001; 2001US-0271839P.  
PR 27-FEB-2001; 2001US-0271855P.  
PR 02-MAR-2001; 2001US-0272788P.

PR 02-MAR-2001; 2001US-0273046P.  
 PR 14-MAR-2001; 2001US-0275952P.  
 PR 14-MAR-2001; 2001US-0275947P.  
 PR 14-MAR-2001; 2001US-0275950P.  
 PR 14-MAR-2001; 2001US-0275989P.  
 PR 15-MAR-2001; 2001US-0276448P.  
 PR 15-MAR-2001; 2001US-0276450P.  
 PR 16-MAR-2001; 2001US-0276397P.  
 PR 16-MAR-2001; 2001US-0276768P.  
 PR 20-MAR-2001; 2001US-0278652P.  
 PR 26-MAR-2001; 2001US-0278752P.  
 PR 26-MAR-2001; 2001US-0278778P.  
 PR 29-MAR-2001; 2001US-0279882P.  
 PR 29-MAR-2001; 2001US-0279884P.  
 PR 30-MAR-2001; 2001US-0280147P.  
 PR 11-APR-2001; 2001US-0282922P.  
 PR 11-APR-2001; 2001US-0283083P.  
 PR 20-APR-2001; 2001US-0285133P.  
 PR 23-APR-2001; 2001US-0285749P.  
 PR 03-MAY-2001; 2001US-0288327P.  
 PR 03-MAY-2001; 2001US-0288504P.  
 PR 29-MAY-2001; 2001US-0294047P.  
 PR 30-MAY-2001; 2001US-0294473P.  
 PR 08-JUN-2001; 2001US-0296964P.  
 PR 18-JUN-2001; 2001US-0298959P.  
 PR 19-JUN-2001; 2001US-0299324P.  
 PR 13-AUG-2001; 2001US-0312020P.  
 PR 16-AUG-2001; 2001US-0312889P.  
 PR 16-AUG-2001; 2001US-0312908P.  
 PR 21-AUG-2001; 2001US-0313390P.  
 PR 28-AUG-2001; 2001US-0315470P.  
 PR 31-AUG-2001; 2001US-0316447P.  
 PR 07-SEP-2001; 2001US-0318115P.  
 PR 07-SEP-2001; 2001US-0318118P.  
 PR 12-SEP-2001; 2001US-0318740P.  
 PR 19-SEP-2001; 2001US-0323379P.  
 PR 18-OCT-2001; 2001US-0330245P.  
 PR 18-OCT-2001; 2001US-0330308P.  
 PR 14-NOV-2001; 2001US-0332701P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Tchernev VT, Solytek KA, Zehrhusen BD, Patnurejan M, Shinkete RA;  
 PI Li L, Gangolli EA, Padigam M, Anderson DM, Rastelli L, Miller CE;  
 PI Gerlach VN, Taupier RJ, Gusev VY, Colman SD, Wolenc AK, Pena CE;  
 PI Futak K, Grosse WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;  
 XX  
 XX WPI; 2002-706998/76.  
 DR  
 XX  
 PT New NOVX polypeptides and nucleic acids, useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 PT  
 XX  
 PS Disclosure; SEQ ID NO 419; 1498bp; English.  
 XX  
 CC This invention relates to a novel nucleic acid, and encoded polypeptides  
 CC thereof, which have properties related to the stimulation of biochemical  
 CC or physiological responses in a cell, tissue, organ or organism.  
 CC Specifically, it refers to the use of biologically active fragments for  
 CC diagnostic and prognostic assays and furthermore in the treatment of  
 CC diverse pathological conditions. The present invention describes novel  
 CC human and murine NOVX proteins, as well as methods to modulate their  
 CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
 CC The NOVX polypeptides, polynucleotides and antibodies are useful in  
 CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,  
 CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
 CC treating or preventing diseases such as inflammation, autoimmune  
 CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
 CC (AIDS), obesity, asthma, immunoglobulin (IgA) nephropathy, cirrhosis,  
 CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
 CC and epilepsy. Accordingly, these molecules have many activities including  
 CC cyostatic, cardant, antiinflammatory, immunosuppressive, antiallergic,

CC haemostatic, anti-HIV, antidiabetic, antiatherosclerotic, anorectic,  
 CC antiaesthetic, nephrotoxic, antihypertensive, hepatotropic,  
 CC neuroprotective, nootropic, antibacterial, virucide, antiparasitic,  
 CC relaxant and anticoagulant. In addition, they are useful in screening  
 CC assays to identify small molecules that modulate or inhibit, for example,  
 CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
 CC used as in chromosome mapping, tissue typing, preventive medicine and  
 CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein  
 CC of the invention.  
 XX  
 SQ Sequence 855 AA;  
 Query Match 99.2%; Score 3760; DB 5; Length 855;  
 Best Local Similarity 99.4%; Pred. No. 4.7e-241;  
 Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MAERVVMLEPPARSLKSFVVTSSVVAEPDTSKTQRTQDSCSPGLHARGVELMRFTTPG 60  
 DB 173 MAERVVMLEPPARSLKSFVVTSSVVAEPDTSKTQRTQDSCSPGLHARGVELMRFTTPG 232  
 QY 61 PPDSPPYPAHARCGWALGADSVSLTFRSPDLASCBERGSDLVTVYNTLSPMEPHALVQ 120  
 DB 233 PPDSPPYPAHARCGWALGADSVSLTFRSPDLASCBERGSDLVTVYNTLSPMEPHALVQ 292  
 QY 121 LCGTTPPSVNLTFHSSQWVLLITLITNTRRHHPGEATFPOLPRMSSCGRLRAQGTN 180  
 DB 293 LCGTTPPSVNLTFHSSQWVLLITLITNTRRHHPGEATFPOLPRMSSCGRLRAQGTN 352  
 QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVVRKFFYLLEPRRACGTCPKDYVEINERKTCGE 240  
 DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVVRKFFYLLEPRRACGTCPKDYVEINERKTCGE 412  
 QY 241 RSQFVVTNSNKTIVRHSQSDSYTDFTFLAEYLSYDSDPCPGQFTRTGCIKELKCD 300  
 DB 413 RSQFVVTNSNKTIVRHSQSDSYTDFTFLAEYLSYDSDPCPGQFTRTGCIKELKCD 472  
 QY 301 GMADCTHSDDELNCSGAGHOFCTCKNFKCPLEFVNCDSVNDCCGNSDEGSCCPAQTFRC 360  
 DB 473 GMADCTHSDDELNCSGAGHOFCTCKNFKCPLEFVNCDSVNDCCGNSDEGSCCPAQTFRC 532  
 QY 361 SNGKLSKSOQCGKXGDCGSDSEASCPKVVVYCTKHTYRCNLGLCLSKGNECDKED 420  
 DB 533 SNGKLSKSOQCGKXGDCGSDSEASCPKVVVYCTKHTYRCNLGLCLSKGNECDKED 592  
 QY 421 CSDGSDKDCDCCGRLSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLSPWLV 480  
 DB 593 CSDGSDKDCDCCGRLSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLSPWLV 652  
 QY 481 SAHCYIDDRGFRSDPTQWTAFLGLHDOSORSAFGVORRLKRIISHPFNDTFPYDI 540  
 DB 653 SAHCYIDDRGFRSDPTQWTAFLGLHDOSORSAFGVORRLKRIISHPFNDTFPYDI 712  
 QY 541 ALLELEPAEYSSVWRPICLPDASHVPAGKAIWVTMGHTQYGGTALLLOKEIRVIN 600  
 DB 713 ALLELEPAEYSSVWRPICLPDASHVPAGKAIWVTMGHTQYGGTALLLOKEIRVIN 772  
 QY 601 QTTCENTLLPQITPRMNCVGFSLGCVNCCGDSGGLPSVADGRITQAGVSWGDCAQ 660  
 DB 773 QTTCENTLLPQITPRMNCVGFSLGCVNCCGDSGGLPSVADGRITQAGVSWGDCAQ 832  
 QY 661 RNKRGVYTRLPFLPDWIKENTGV 683  
 DB 833 RNKRGVYTRLPFLPDWIKENTGV 855  
 RESULT 7  
 AD116876  
 XX AD116876 standard; protein; 855 AA.  
 XX  
 AC AD116876;  
 XX  
 DT 15-APR-2004 (first entry)  
 XX

DE Human NOVX protein homologue SegID 412.  
XX human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;  
XX inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; str.  
XX Homo sapiens.  
XX  
XX WO200268649-A2.  
XX  
XX 06-SEP-2002.  
XX  
XX 31-JAN-2002; 2002WO-US002785.  
XX  
XX 31-JAN-2001; 2001US-0265395P.  
XX 31-JAN-2001; 2001US-0265412P.  
XX 31-JAN-2001; 2001US-0265514P.  
XX 31-JAN-2001; 2001US-0265517P.  
XX 02-FEB-2001; 2001US-0266406P.  
XX 05-FEB-2001; 2001US-0266767P.  
XX 07-FEB-2001; 2001US-0266975P.  
XX 07-FEB-2001; 2001US-0267057P.  
XX 08-FEB-2001; 2001US-0267459P.  
XX 09-FEB-2001; 2001US-0267823P.  
XX 15-FEB-2001; 2001US-0268974P.  
XX 26-FEB-2001; 2001US-0271664P.  
XX 27-FEB-2001; 2001US-0271839P.  
XX 27-FEB-2001; 2001US-0271855P.  
XX 02-MAR-2001; 2001US-0272788P.  
XX 02-MAR-2001; 2001US-0273046P.  
XX 14-MAR-2001; 2001US-0275925P.  
XX 14-MAR-2001; 2001US-0275947P.  
XX 14-MAR-2001; 2001US-0275950P.  
XX 14-MAR-2001; 2001US-0275989P.  
XX 15-MAR-2001; 2001US-0276448P.  
XX 15-MAR-2001; 2001US-0276450P.  
XX 16-MAR-2001; 2001US-0276397P.  
XX 16-MAR-2001; 2001US-0276768P.  
XX 20-MAR-2001; 2001US-0276852P.  
XX 26-MAR-2001; 2001US-0278778P.  
XX 26-MAR-2001; 2001US-0278778P.  
XX 29-MAR-2001; 2001US-0279882P.  
XX 30-MAR-2001; 2001US-0279884P.  
XX 11-APR-2001; 2001US-0280147P.  
XX 11-APR-2001; 2001US-0282992P.  
XX 20-APR-2001; 2001US-0283083P.  
XX 20-APR-2001; 2001US-0285133P.  
XX 23-APR-2001; 2001US-0285749P.  
XX 03-MAY-2001; 2001US-0288327P.  
XX 03-MAY-2001; 2001US-0288504P.  
XX 29-MAY-2001; 2001US-0294047P.  
XX 30-MAY-2001; 2001US-0294473P.  
XX 08-JUN-2001; 2001US-0296964P.  
XX 18-JUN-2001; 2001US-0298959P.  
XX 19-JUN-2001; 2001US-0299324P.  
XX 13-AUG-2001; 2001US-0312020P.  
XX 15-AUG-2001; 2001US-0312889P.  
XX 16-AUG-2001; 2001US-0312908P.  
XX 21-AUG-2001; 2001US-0313390P.  
XX 28-AUG-2001; 2001US-0316470P.  
XX 31-AUG-2001; 2001US-0316470P.  
XX 07-SEP-2001; 2001US-0318115P.  
XX 07-SEP-2001; 2001US-0318118P.  
XX 12-SEP-2001; 2001US-0318740P.  
XX 19-SEP-2001; 2001US-0323379P.  
XX 18-OCT-2001; 2001US-0330245P.  
XX 18-OCT-2001; 2001US-0330308P.  
XX 14-NOV-2001; 2001US-0332701P.  
XX  
XX (CURA-) CURAGEN CORP.

PI Therneer VT, Spytek KA, Zernusen BD, Paturajan M, Shinkens RA;  
PI Li L, Gangoli EA, Padigaru M, Anderson DW, Rastelli L, Miller CE;  
PI Gerlach U, Taupier JR, Gusev VV, Colman SD, Wolenc AR, Pena CE;  
PI Furtak K, Grosse WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;  
XX WPI, 2002-706998/76.  
DR  
XX  
XX  
XX New NOVX polypeptides and nucleic acids, useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX  
XX  
XX Disclosure; SEQ ID NO 412; 1498bp; English.  
XX  
XX  
XX This invention relates to a novel nucleic acids, and encoded polypeptides  
CC thereof, which have properties related to the stimulation of biochemical  
CC or physiological responses in a cell, tissue, organ or organism.  
CC Specifically, it refers to the use of biologically active fragments for  
CC diagnostic and prognostic assays and furthermore in the treatment of  
CC diverse pathological conditions. The present invention describes novel  
CC human and murine NOVX proteins, as well as methods to modulate their  
CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
CC The NOVX polypeptides, polynucleotides and antibodies are useful in  
CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,  
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
CC treating or preventing diseases such as inflammation, autoimmune  
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,  
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
CC and epilepsy. Accordingly, these molecules have many activities including  
CC cyostatic, cardiac, anti-inflammatory, immunosuppressive, antiallergic,  
CC hemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,  
CC antisthmatic, nephrotoxic, antibacterial, virucide, antiparasitic,  
CC neuroprotective, nootropic, anticholinergic, antipruritic, antipruritic,  
CC relaxant and anticonvulsant. In addition, they are useful in screening  
CC assays to identify small molecules that modulate or inhibit, for example,  
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
CC used as in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein  
CC of the invention.  
XX  
XX  
SQ Sequence 855 AA,  
Query Match 99.2%; Score 3760; DB 5; Length 855;  
Best local Similarity 99.4%; Pred. No. 4,7e-241;  
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MAERVYMLPPRARSLKSPFVYTSVVAFPDTSKVORTQDSCSGFGLHARGVELMRFTTPG 60  
DB 173 MAERVYMLPPRARSLKSPFVYTSVVAFPDTSKVORTQDSCSGFGLHARGVELMRFTTPG 232  
QY 61 FPDSPYPAHARCMALRGDADSVLSTFPRSDIASCDERSGDLVTYNTLSPMEHALVQ 120  
DB 233 FPDSPYPAHARCMALRGDADSVLSTFPRSDIASCDERSGDLVTYNTLSPMEHALVQ 232  
QY 121 LCGTYPSPSYNLTFHSSQNVLLTLINTERRHPGPATFQPLFRMSSCGRLAKAQTEN 180  
DB 293 LCGTYPSPSYNLTFHSSQNVLLTLINTERRHPGPATFQPLFRMSSCGRLAKAQTEN 352  
QY 181 SPYYPGHYPNIDCTNIEVPNNQHVAFKFPYLLPRAAGCTCPKDYVINGEYXCGE 240  
DB 353 SPYYPGHYPNIDCTNIEVPNNQHVAFKFPYLLPRAAGCTCPKDYVINGEYXCGE 412  
QY 241 RSQFVYTSNSNKTITVAFHSDQSYTDGFLAELYSYDSDPCPGQFTCRGRCIRKEIRCD 300  
DB 413 RSQFVYTSNSNKTITVAFHSDQSYTDGFLAELYSYDSDPCPGQFTCRGRCIRKEIRCD 472  
QY 301 GWADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVCDSVNDGSDNDEGSCCPAOTFRC 360  
DB 473 GWADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVCDSVNDGSDNDEGSCCPAOTFRC 532  
QY 361 SNGCKLSKSGQCGKDCGDSGDEASCPRVNVVYTCRHYRCINGLCLSGNDEGCKED 420

Db 533 SNGKCLSKSQCKNGKDDCGSDDEASCPKVVVVTCTKHYRCLNGLCLSKGNPECDKED 592  
Qy 421 CSDSDSDKDCCGCLRSFTFRQARVVGCTADADGEMPMOVSJLALQGGHICGASLISPMMLV 480  
Db 593 CSDSDSDKDCCGCLRSFTFRQARVVGCTADADGEMPMOVSJLALQGGHICGASLISPMMLV 652  
Qy 481 SAACGYIDDRGFRYSPTQMTAFGLHDSORSAPGVGERLKRILISHPFNDFTFDYDI 540  
Db 653 SAACGYIDDRGFRYSPTQMTAFGLHDSORSAPGVGERLKRILISHPFNDFTFDYDI 712  
Qy 541 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVGWGTGYGGTGAIILOKGEIRVIN 600  
Db 713 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVGWGTGYGGTGAIILOKGEIRVIN 772  
Qy 601 OTTGENLLPQGITRRMVCVGLSGVDSGQSDSGPLSSVADGRIFPAGVISMDCGAQ 660  
Db 773 QTTEENLLPQGITRRMVCVGLSGVDSGQSDSGPLSSVADGRIFPAGVISMDCGAQ 832  
Qy 661 RNKGVYTRLPFRDWMKENTGV 683  
Db 833 RNKGVYTRLPFRDWMKENTGV 855

RESULT 8  
ADN39867  
ID ADN39867 standard; protein; 855 AA.  
XX  
AC ADN39867;  
XX  
DT 17-JUN-2004 (first entry)  
XX  
DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO: C237.  
XX  
KW Human; differential expression; cancer; angiogenic disorder;  
KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
KW inflammatory disease; autoimmune disease;  
KW retinal neovascularization syndrome; scarring; uterine fibroid;  
KW detection; diagnosis; prognosis; drug screening; drug targeting;  
KW wound healing; contraception; cytostatic; cardiac; immunomodulatory;  
KW vulnery; gene therapy; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN W02003042661-A2.  
XX  
PD 22-MAY-2003.  
XX  
PF 13-NOV-2002; 2002WO-US036810.  
XX  
PR 13-NOV-2001; 2001US-0350666P.  
PR 21-NOV-2001; 2001US-0332464P.  
PR 29-NOV-2001; 2001US-0334393P.  
PR 03-DEC-2001; 2001US-0353394P.  
PR 14-DEC-2001; 2001US-0340376P.  
PR 08-JAN-2002; 2002US-0347211P.  
PR 10-JAN-2002; 2002US-0347349P.  
PR 08-FEB-2002; 2002US-0355250P.  
PR 13-FEB-2002; 2002US-0356714P.  
PR 20-FEB-2002; 2002US-0359077P.  
PR 29-MAR-2002; 2002US-0368080P.  
PR 04-APR-2002; 2002US-0370110P.  
PR 12-APR-2002; 2002US-0372246P.  
PR 05-JUN-2002; 2002US-0386614P.  
PR 16-JUL-2002; 2002US-0396839P.  
PR 22-JUL-2002; 2002US-039775P.  
PR 22-JUL-2002; 2002US-0397845P.  
PR 09-SEP-2002; 2002US-0409450P.  
XX  
PA (EOSB-) EOS BIOTECHNOLOGY INC.  
XX  
PI Afeir D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;  
PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;

DR WPI: 2003-468649/44.  
DR N-PSDB; ADN39650.  
XX  
PT Determining the presence or absence of a pathological cell in a patient,  
PT useful for diagnosing, prognosing or treating cancer, comprises detecting  
PT a nucleic acid in a biological sample.  
XX  
PS Claim 12; SEQ ID NO C237; 1385pp; English.  
XX  
CC The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
CC whose expression is upregulated or downregulated in specific cancers or  
CC other diseases such as angiogenic or fibrotic disorders, and to methods  
CC of determining the presence or absence of a pathological cell in a  
CC patient by detecting a nucleic acid at least 80% identical to those of  
CC the invention or by detecting a polypeptide of the invention. The  
CC invention also relates to expression vectors and host cells comprising a  
CC nucleic acid of the invention; antibodies which specifically bind a  
CC polypeptide of the invention; use of such antibodies for drug targeting;  
CC and methods of screening for modulators of activity or expression of the  
CC polypeptides and nucleic acids. The nucleic acids, polypeptides,  
CC antibodies and methods are useful for diagnosing, prognosing and treating  
CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
CC neovascularization syndromes, scarring and uterine fibroids. They may  
CC also be useful in wound healing and in contraception. The present  
CC sequence represents a polypeptide of the invention.  
XX  
SQ Sequence 855 AA;  
Query Match 99.2%; Score 3760; DB 7; Length 855;  
Best Local Similarity 99.4%; Pred. No. 4,7e-241;  
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 MAERVVMLPPRARSLKSFVVTYSVAPPTSKYQRTQDNSCSFGHARVGLMRFTTPG 60  
Db 173 MAERVVMLPPRARSLKSFVVTYSVAPPTSKYQRTQDNSCSFGHARVGLMRFTTPG 232  
Qy 61 PPDSPPYPAHARCOMALGDDSVSLTFRRSFDLASCEBRSDDVTVYNTLSPMEPHALVQ 120  
Db 233 PPDSPPYPAHARCOMALGDDSVSLTFRRSFDLASCEBRSDDVTVYNTLSPMEPHALVQ 292  
Qy 121 LCGTYPPSYNLTFHSSQNVLLITLITNTERRHQFEATFPOLPMSCGRLRKAQSTFN 180  
Db 293 LCGTYPPSYNLTFHSSQNVLLITLITNTERRHQFEATFPOLPMSCGRLRKAQSTFN 352  
Qy 181 SPYYPGHYPNIDCTWNIIEVPNNQHVVRKFFYLLEPRACGTCPPDYVEINSEKTCGE 240  
Db 353 SPYYPGHYPNIDCTWNIIEVPNNQHVVRKFFYLLEPRACGTCPPDYVEINSEKTCGE 412  
Qy 241 RSQFVVTNSNKTIVRHSOSYTDGFLAEVLSYSDSDCPQGFCTORTGCTRKELRCD 300  
Db 413 RSQFVVTNSNKTIVRHSOSYTDGFLAEVLSYSDSDCPQGFCTORTGCTRKELRCD 472  
Qy 301 GWADCTHSDDELNCSGAGHQFTCKNFKCPLFWVCDSVANDCGNSDEGCSGPAQTRFC 360  
Db 473 GWADCTHSDDELNCSGAGHQFTCKNFKCPLFWVCDSVANDCGNSDEGCSGPAQTRFC 532  
Qy 361 SNGKCLSKSQCKNGKDDCGSDDEASCPKVVVVTCTKHYRCLNGLCLSKGNPECDKED 420  
Db 533 SNGKCLSKSQCKNGKDDCGSDDEASCPKVVVVTCTKHYRCLNGLCLSKGNPECDKED 592  
Qy 421 CSDSDSDKDCCGCLRSFTFRQARVVGCTADADGEMPMOVSJLALQGGHICGASLISPMMLV 480  
Db 593 CSDSDSDKDCCGCLRSFTFRQARVVGCTADADGEMPMOVSJLALQGGHICGASLISPMMLV 652  
Qy 481 SAACGYIDDRGFRYSPTQMTAFGLHDSORSAPGVGERLKRILISHPFNDFTFDYDI 540  
Db 653 SAACGYIDDRGFRYSPTQMTAFGLHDSORSAPGVGERLKRILISHPFNDFTFDYDI 712  
Qy 541 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVGWGTGYGGTGAIILOKGEIRVIN 600  
Db 713 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVGWGTGYGGTGAIILOKGEIRVIN 772

QY 601 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSCGSPLSVEADGRIFQAGVSWGDCAQ 660  
DB 773 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSCGSPLSVEADGRIFQAGVSWGDCAQ 832  
QY 661 RNRKGVYTRLPFRDWIKENTGV 683  
DB 833 RNRKGVYTRLPFRDWIKENTGV 855

## RESULT 9

ADN04754  
ID ADN04754 standard; protein; 855 AA.

AC ADN04754;

DT 01-JUL-2004 (first entry)

DE Antipsoriatic protein sequence #558.

XX antipsoriatic; gene therapy; psoriasis; diagnosis.

OS Homo sapiens.

XX WO2004028479-A2.

XX 08-APR-2004.

PF 25-SEP-2003; 2003WO-US030907.

PR 25-SEP-2002; 2002US-0414006P.

PA (GETH ) GENENTECH INC.

PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WJ,  
Wu TD;

DR WPI; 2004-305105/28.

DR N-PSDB; ADN04753.

PT New PRO nucleic acid or polypeptide, useful for preparing a  
pharmaceutical composition for diagnosing or treating psoriasis in a  
mammal.

XX Claim 9; SEQ ID NO 1148; 3063pp; English.

XX The invention relates to novel polynucleotide and polypeptides for  
treating psoriasis or a sequence having at least 80% identity to the  
above sequences. The nucleic acid is useful for preparing a composition  
for diagnosing or treating psoriasis in a mammal. This sequence  
corresponds to one of the polypeptides of the invention.

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

QY 241 RSQFVYVTSNKNKITTFRHSDQSYTDGFLAEVLSYDSDBCPQOFTCRTRCIRKEIRCD 300  
DB 413 RSQFVYVTSNKNKITTFRHSDQSYTDGFLAEVLSYDSDBCPQOFTCRTRCIRKEIRCD 472  
QY 301 GMAADCTDHSDELNCSQDAGHQTCKNKFCKPLFWVCDVNDCCDSDNDEQSCCPAOTFRC 360  
DB 473 GMAADCTDHSDELNCSQDAGHQTCKNKFCKPLFWVCDVNDCCDSDNDEQSCCPAOTFRC 532  
QY 361 SNGKCLSKSQCGKNGKDCDGDSDASCPKVVVTCRTKTYRCINGLCLSKGNPECDKED 420  
DB 533 SNGKCLSKSQCGKNGKDCDGDSDASCPKVVVTCRTKTYRCINGLCLSKGNPECDKED 592  
QY 421 CSDGSEKDCDGLRSPTRQARVVGTDADGEWPMQVSHALGQGHICGASLISPMWLV 480  
DB 593 CSDGSEKDCDGLRSPTRQARVVGTDADGEWPMQVSHALGQGHICGASLISPMWLV 652  
QY 481 SAACHYIDDRGFRYSPTQWTAFLGLHDSQSRAPGVQERLRRIISHPEFNDFTDYDI 540  
DB 653 SAACHYIDDRGFRYSPTQWTAFLGLHDSQSRAPGVQERLRRIISHPEFNDFTDYDI 712  
QY 541 ALLELEKPAEYSSWBPICLPDASHVFPAGKAIWVTGMGHTGYGAGALILQKGEIRVIN 600  
DB 713 ALLELEKPAEYSSWBPICLPDASHVFPAGKAIWVTGMGHTGYGAGALILQKGEIRVIN 772  
QY 601 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSCGSPLSVEADGRIFQAGVSWGDCAQ 660  
DB 773 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSCGSPLSVEADGRIFQAGVSWGDCAQ 832  
QY 661 RNRKGVYTRLPFRDWIKENTGV 683  
DB 833 RNRKGVYTRLPFRDWIKENTGV 855

## RESULT 10

AD116884  
ID AD116884 standard; protein; 855 AA.

AC AD116884;

DT 15-APR-2004 (first entry)

DE Human NOXV protein homologue Segid 420.

XX human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;  
inflammation; autoimmune disorder; allergy; blood disorder;  
acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
Alzheimer's disease; infection; str.

XX Homo sapiens.

XX WO200268649-A2.

XX 06-SEP-2002.

PF 31-JAN-2002; 2002WO-US002785.

PA 31-JAN-2001; 2001US-0265395P.

PI 31-JAN-2001; 2001US-0265412P.

DR 31-JAN-2001; 2001US-0265517P.

DR 31-JAN-2001; 2001US-0265517P.

DR 02-FEB-2001; 2001US-0266406P.

DR 05-FEB-2001; 2001US-0266767P.

DR 07-FEB-2001; 2001US-0266757P.

DR 07-FEB-2001; 2001US-0267057P.

DR 08-FEB-2001; 2001US-0267459P.

DR 09-FEB-2001; 2001US-0267823P.

DR 15-FEB-2001; 2001US-0268974P.

DR 26-FEB-2001; 2001US-0271664P.

DR 27-FEB-2001; 2001US-0271839P.

DR 27-FEB-2001; 2001US-0271855P.

DR 02-MAR-2001; 2001US-0272788P.

DR 02-MAR-2001; 2001US-0273046P.

PR 14-MAR-2001; 2001US-0275925P.  
PR 14-MAR-2001; 2001US-0275947P.  
PR 14-MAR-2001; 2001US-0275950P.  
PR 14-MAR-2001; 2001US-0275989P.  
PR 15-MAR-2001; 2001US-0276448P.  
PR 15-MAR-2001; 2001US-0276450P.  
PR 16-MAR-2001; 2001US-0276397P.  
PR 16-MAR-2001; 2001US-0276768P.  
PR 20-MAR-2001; 2001US-0276852P.  
PR 26-MAR-2001; 2001US-0278775P.  
PR 26-MAR-2001; 2001US-0278778P.  
PR 29-MAR-2001; 2001US-0279882P.  
PR 29-MAR-2001; 2001US-0279884P.  
PR 30-MAR-2001; 2001US-0280147P.  
PR 11-APR-2001; 2001US-0282992P.  
PR 11-APR-2001; 2001US-0283083P.  
PR 20-APR-2001; 2001US-0285133P.  
PR 23-APR-2001; 2001US-0285749P.  
PR 03-MAY-2001; 2001US-0288327P.  
PR 03-MAY-2001; 2001US-0288504P.  
PR 29-MAY-2001; 2001US-0294047P.  
PR 30-MAY-2001; 2001US-0294473P.  
PR 08-JUN-2001; 2001US-0296964P.  
PR 18-JUN-2001; 2001US-0298959P.  
PR 19-JUN-2001; 2001US-0299324P.  
PR 13-AUG-2001; 2001US-0312020P.  
PR 16-AUG-2001; 2001US-0312889P.  
PR 16-AUG-2001; 2001US-0312908P.  
PR 21-AUG-2001; 2001US-0313390P.  
PR 28-AUG-2001; 2001US-0315470P.  
PR 31-AUG-2001; 2001US-0316447P.  
PR 07-SEP-2001; 2001US-0318115P.  
PR 07-SEP-2001; 2001US-0318118P.  
PR 12-SEP-2001; 2001US-0318740P.  
PR 19-SEP-2001; 2001US-0323379P.  
PR 18-OCT-2001; 2001US-0330245P.  
PR 18-OCT-2001; 2001US-0330308P.  
PR 14-NOV-2001; 2001US-0332701P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX Tchernev VT, Spytek KA, Zernhusen BD, Paturajan M, Shinkets RA;  
PI Li L, Gargolli EA, Padigaru M, Anderson DM, Rastelli L, Miller CB;  
PI Gelach VL, Taudier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CE;  
PI Putrak K, Groese WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;  
XX WPI; 2002-706998/76.  
XX  
XX New NOXV polypeptides and nucleic acids, useful for preventing or  
PT treating NOXV-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX  
XX Disclosure; SEQ ID NO 420; 1498bp; English.  
XX  
XX This invention relates to a novel nucleic acid, and encoded polypeptides  
CC thereof, which have properties related to the stimulation of biochemical  
CC or physiological responses in a cell, tissue, organ or organism.  
CC Specifically, it refers to the use of biologically active fragments for  
CC diagnostic and prognostic assays and furthermore in the treatment of  
CC diverse pathological conditions. The present invention describes novel  
CC human and murine NOXV proteins, as well as methods to modulate their  
CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
CC The NOXV polypeptides, polynucleotides and antibodies are useful in  
CC treating or preventing NOXV-associated disorders, e.g. cardiomyopathy,  
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
CC treating or preventing diseases such as inflammation, autoimmune  
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,  
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
CC and epilepsy. Accordingly, these molecules have many activities including  
CC cytostatic, cardiac, antiinflammatory, immunosuppressive, antiallergic,  
CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,

CC antiasthmatic, nephrotropic, antiarthritic, hepatotropic,  
CC neuroprotective, nootropic, antibacterial, virucide, antiparasitic,  
CC relaxant and anticonvulsant. In addition, they are useful in screening  
CC assays to identify small molecules that modulate or inhibit, for example,  
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
CC used as in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. This polypeptide is a homologue of a human NOXV protein  
CC of the invention.  
XX  
XX Sequence 855 AA;  
SQ  
Query Match 99.1%; Score 3756; DB 5; Length 855;  
Best Local Similarity 99.3%; Pred. No. 8,6e-241;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MAERVVMLEPPRASLSFVVTSSVVAPEPTDSTKIQORQDNSSCGFLHARGVELMRFPTPG 60  
DB 173 MAERVVMLEPPRASLSFVVTSSVVAPEPTDSTKIQORQDNSSCGFLHARGVELMRFPTPG 232  
QY 61 PPDSPPYPAHARCOMALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMEPALVQ 120  
DB 233 PPDSPPYPAHARCOMALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMEPALVQ 292  
QY 121 LCGTYPSPSYNLTFSSQNVLLITLITNTERRHHPFEATFFQLPRMSSCGRLRAQGTFN 180  
DB 293 LCGTYPSPSYNLTFSSQNVLLITLITNTERRHHPFEATFFQLPRMSSCGRLRAQGTFN 352  
QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVYRFFEFLLBPRACGTCPPDYVINEKTCGE 240  
DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVYRFFEFLLBPRACGTCPPDYVINEKTCGE 412  
QY 241 RSQFVMTNSNKKIVRFRHSQSYDTDTGFLAEYLSYDSDDPCPGQFTCRGCIKELRCD 300  
DB 413 RSQFVMTNSNKKIVRFRHSQSYDTDTGFLAEYLSYDSDDPCPGQFTCRGCIKELRCD 472  
QY 301 GMAOCTHSDDELNCSCDAGHOFCKNFKCKPLFVVCDSVNDGDNSSDQSCSPAQTFC 360  
DB 473 GMAOCTHSDDELNCSCDAGHOFCKNFKCKPLFVVCDSVNDGDNSSDQSCSPAQTFC 532  
QY 361 SNGKCLSKSQCKGCKDCCGSDSDASCPKVVNTCTGTYTRCLNGLCLSKNPECDKED 420  
DB 533 SNGKCLSKSQCKGCKDCCGSDSDASCPKVVNTCTGTYTRCLNGLCLSKNPECDKED 592  
QY 421 CSDSDSDKDCCGRSTRTRARVYVGTADGEMPMQVSLHAGQGHICGASLISPMMLV 480  
DB 593 CSDSDSDKDCCGRSTRTRARVYVGTADGEMPMQVSLHAGQGHICGASLISPMMLV 652  
QY 481 SAACHYIDRGFRYSDEPTOWTAFLGLHDQSRASAPGQERRLRKRIISHPFENDETFPDYDI 540  
DB 653 SAACHYIDRGFRYSDEPTOWTAFLGLHDQSRASAPGQERRLRKRIISHPFENDETFPDYDI 712  
QY 541 ALLELEKPAEYSSWVRPCLCPDASHVPAGKAIWVTGWHGTQYGGTALLIQKEIRVIN 600  
DB 713 ALLELEKPAEYSSWVRPCLCPDASHVPAGKAIWVTGWHGTQYGGTALLIQKEIRVIN 772  
QY 601 OTTCENLIPQITRRMNCVGTGLSGVNSCGDSCGSPSYEADRIROAGVNSGDDCAQ 660  
DB 773 OTTCENLIPQITRRMNCVGTGLSGVNSCGDSCGSPSYEADRIROAGVNSGDDCAQ 832  
QY 661 RNKPGVYTRLPFLFDMWKENTGV 683  
DB 833 RNKPGVYTRLPFLFDMWKENTGV 855  
RESULT 11  
AD116818  
ID AD116818 standard; protein; 855 AA.  
XX  
XX AD116818;  
XX  
XX 15-APR-2004 (first entry)  
XX  
XX Human NOXV protein homologue Segid 354.

XX human; NOX; cardiomyopathy; atherosclerosis; cancer; diabetes;  
KW inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; str.  
XX Homo sapiens.  
XX  
XX WO200268649-A2.  
XX  
XX 06-SEP-2002.  
XX  
XX 31-JAN-2002; 2002WO-US002785.  
XX  
XX 31-JAN-2001; 2001US-0265395P.  
XX 31-JAN-2001; 2001US-0265412P.  
XX 31-JAN-2001; 2001US-0265514P.  
XX 31-JAN-2001; 2001US-0265517P.  
XX 02-FEB-2001; 2001US-0266406P.  
XX 05-FEB-2001; 2001US-0266767P.  
XX 07-FEB-2001; 2001US-0266975P.  
XX 07-FEB-2001; 2001US-0267057P.  
XX 08-FEB-2001; 2001US-0267459P.  
XX 09-FEB-2001; 2001US-0267823P.  
XX 15-FEB-2001; 2001US-0268974P.  
XX 26-FEB-2001; 2001US-0271664P.  
XX 27-FEB-2001; 2001US-0271839P.  
XX 27-FEB-2001; 2001US-0271855P.  
XX 02-MAR-2001; 2001US-0272788P.  
XX 02-MAR-2001; 2001US-0273046P.  
XX 14-MAR-2001; 2001US-0275925P.  
XX 14-MAR-2001; 2001US-0275947P.  
XX 14-MAR-2001; 2001US-0275950P.  
XX 14-MAR-2001; 2001US-0275989P.  
XX 15-MAR-2001; 2001US-0276448P.  
XX 15-MAR-2001; 2001US-0276450P.  
XX 16-MAR-2001; 2001US-0276379P.  
XX 16-MAR-2001; 2001US-0276768P.  
XX 20-MAR-2001; 2001US-0278652P.  
XX 26-MAR-2001; 2001US-0278775P.  
XX 26-MAR-2001; 2001US-0278778P.  
XX 29-MAR-2001; 2001US-0279862P.  
XX 29-MAR-2001; 2001US-0279884P.  
XX 30-MAR-2001; 2001US-0280147P.  
XX 11-APR-2001; 2001US-0282992P.  
XX 11-APR-2001; 2001US-0283083P.  
XX 20-APR-2001; 2001US-0285133P.  
XX 23-APR-2001; 2001US-0285749P.  
XX 03-MAY-2001; 2001US-0288347P.  
XX 03-MAY-2001; 2001US-0288504P.  
XX 29-MAY-2001; 2001US-0294047P.  
XX 30-MAY-2001; 2001US-0294473P.  
XX 08-JUN-2001; 2001US-0296964P.  
XX 18-JUN-2001; 2001US-0298959P.  
XX 19-JUN-2001; 2001US-0299324P.  
XX 13-AUG-2001; 2001US-0312020P.  
XX 16-AUG-2001; 2001US-0312889P.  
XX 16-AUG-2001; 2001US-0312908P.  
XX 21-AUG-2001; 2001US-0313390P.  
XX 28-AUG-2001; 2001US-0313470P.  
XX 31-AUG-2001; 2001US-0316447P.  
XX 07-SEP-2001; 2001US-0318115P.  
XX 07-SEP-2001; 2001US-0318118P.  
XX 12-SEP-2001; 2001US-0318740P.  
XX 19-SEP-2001; 2001US-0323379P.  
XX 18-OCT-2001; 2001US-0330245P.  
XX 18-OCT-2001; 2001US-0330308P.  
XX 14-NOV-2001; 2001US-0332701P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Tchernev VT, Spytek KA, Zethtusen BD, Patturajan M, Shimkets RA;

PI Li L, Gangoli EA, Padigar M, Anderson DW, Rastelli L, Miller CE,  
PI Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CBA,  
PI Futrak K, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK, Burgess CF;  
XX WPI; 2002-706998/76.  
XX  
XX  
XX New NOX polypeptides and nucleic acids, useful for preventing or  
PT treating NOX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX  
XX  
XX Disclosure, SEQ ID NO 354; 1498pp; English.  
XX  
XX  
XX This invention relates to a novel nucleic acids, and encoded polypeptides  
CC thereof, which have properties related to the stimulation of biochemical  
CC or physiological responses in a cell, tissue, organ or organism.  
CC Specifically, it refers to the use of biologically active fragments for  
CC diagnostic and prognostic assays and furthermore in the treatment of  
CC diverse pathological conditions. The present invention describes novel  
CC human and murine NOX proteins, as well as methods to modulate their  
CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
CC The NOX polypeptides, polynucleotides and antibodies are useful in  
CC treating or preventing NOX-associated disorders, e.g. cardiomyopathy,  
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
CC treating or preventing diseases such as inflammation, autoimmune  
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,  
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
CC and epilepsy. Accordingly, these molecules have many activities including  
CC cytostatic, cardiant, anti-inflammatory, immunosuppressive, antiallergic,  
CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,  
CC antihistaminic, nephroprotective, antiarthritic, hepatotropic,  
CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,  
CC relaxant and anticoagulant. In addition, they are useful in screening  
CC assays to identify small molecules that modulate or inhibit, for example,  
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
CC used as in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. This polypeptide is a homologue of a human NOX protein  
XX of the invention.  
XX  
SQ Sequence 855 AA;  
Query Match 99.1%; Score 3756; DB 5; Length 855;  
Best Local Similarity 99.3%; Pred. No. 8, 6e-241;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MAERVVM.LPPRARSLKSPFVTSVVA.FPDDSKTVORTQONS.CSGFGLHARGVELMR.FPTTG 60  
DB 173 MAERVVM.LPPRARSLKSPFVTSVVA.FPDDSKTVORTQONS.CSGFGLHARGVELMR.FPTTG 232  
QY 61 FPDSPYPAHAR.CQWALRGDADSVL.SLTFPSFDLAS.CDERGSDLVTVYNTLS.PMEPHALVQ 120  
DB 233 FPDSPYPAHAR.CQWALRGDADSVL.SLTFPSFDLAS.CDERGSDLVTVYNTLS.PMEPHALVQ 232  
QY 121 LCGTTPPSYNTL.TFHSSQNVLL.TLITNTERRHHPGFATTFQ.LPRMSSCGRLKKAQCTRN 180  
DB 293 LCGTTPPSYNTL.TFHSSQNVLL.TLITNTERRHHPGFATTFQ.LPRMSSCGRLKKAQCTRN 352  
QY 181 SPYYPGHYPENIDCTNNIEVNNQHVAFKFPYLL.EPPRACGTCCKDYVEINGEKXCCG 240  
DB 353 SPYYPGHYPENIDCTNNIEVNNQHVAFKFPYLL.EPPRACGTCCKDYVEINGEKXCCG 412  
QY 241 RSQFVVTNSNKT.TVTFHSDOSYTDGFLA.EYLSYDSSDP.CPQOFTCRGTCIRKELRCD 300  
DB 413 RSQFVVTNSNKT.TVTFHSDOSYTDGFLA.EYLSYDSSDP.CPQOFTCRGTCIRKELRCD 472  
QY 301 GWADCTDHSDELNCSCDAGHOF.TCKNFKCPLFWVCD.SVNDGDN.SDBGCSCCPAOTFRG 360  
DB 473 GWADCTDHSDELNCSCDAGHOF.TCKNFKCPLFWVCD.SVNDGDN.SDBGCSCCPAOTFRG 532  
QY 361 SNGKCLSKSQQCNKDCGDCGSD.EASCPRKYNVYTCRKHYYRC.LNGLC.LSKGNPECGCKED 420  
DB 533 SNGKCLSKSQQCNKDCGDCGSD.EASCPRKYNVYTCRKHYYRC.LNGLC.LSKGNPECGCKED 592



QY 421 CSDGSEKDCDCGLRSFTROARVVGCTDADGEMPMOVS LHALGGCHI CGASLISPMNLV 480  
 Db 593 CSDGSEKDCDCGLRSFTROARVVGCTDADGEMPMOVS LHALGGCHI CGASLISPMNLV 652  
 QY 481 SAACHYIDDRGFRYSPTQMTAFGLHDQSQRSAFGVERLKRILSHPFNDFTFYDI 540  
 Db 653 SAACHYIDDRGFRYSPTQMTAFGLHDQSQRSAFGVERLKRILSHPFNDFTFYDI 712  
 QY 541 ALLELEKAEYSNWRPCLPDASHVFPAGKAIWVTGNGHGYGGTGALLIQKEIRVIN 600  
 Db 713 ALLELEKAEYSNWRPCLPDASHVFPAGKAIWVTGNGHGYGGTGALLIQKEIRVIN 772  
 QY 601 OTTCENILPQOITPRMVCVFLSGVDS CQDSSGGLSVADGRIPQAGVSWGDCGQ 660  
 Db 773 OTTCENILPQOITPRMVCVFLSGVDS CQDSSGGLSVADGRIPQAGVSWGDCGQ 832  
 QY 661 RNKPGVYTRLPFRDWIKENTGV 683  
 Db 833 RNKPGVYTRLPFRDWIKENTGV 855

RESULT 12  
 AD116508  
 ID AD116508 standard; protein; 757 AA.  
 XX  
 AC AD116508;  
 XX  
 DT 15-APR-2004 (first entry)  
 XX  
 DE Human NOVX protein to treat human pathological conditions SeqId4.  
 XX  
 KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;  
 KW inflammation; autoimmune disorder; allergy; blood disorder;  
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
 KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
 KW Alzheimer's disease; infection; stroke; muscular dystrophy; epilepsy;  
 KW cytosolic; cardiac; antiinflammatory; immunosuppressive; antiallergic;  
 KW haemostatic; anti-HIV; antidiabetic; antiarteriosclerotic; anorectic;  
 KW antiaesthetic; nephrotoxic; antibacterial; hepatocytotoxic;  
 KW neuroprotective; nootropic; antibacterial; virucide; antiparasitic;  
 KW relaxant; anticonvulsant; neurogenesis; wound healing; angiogenesis;  
 KW chromosome mapping; tissue typing; pharmacogenomic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200268649-A2.  
 XX  
 PD 06-SEP-2002.  
 XX  
 PF 31-JAN-2002; 2002WO-US002785.  
 XX  
 PP 31-JAN-2001; 2001US-0265395P.  
 PR 31-JAN-2001; 2001US-0265412P.  
 PR 31-JAN-2001; 2001US-0265514P.  
 PR 31-JAN-2001; 2001US-0265517P.  
 PR 02-FEB-2001; 2001US-0266406P.  
 PR 05-FEB-2001; 2001US-0266767P.  
 PR 07-FEB-2001; 2001US-0266975P.  
 PR 08-FEB-2001; 2001US-0267057P.  
 PR 09-FEB-2001; 2001US-0267459P.  
 PR 13-FEB-2001; 2001US-0267823P.  
 PR 15-FEB-2001; 2001US-0268974P.  
 PR 26-FEB-2001; 2001US-0271664P.  
 PR 27-FEB-2001; 2001US-0271833P.  
 PR 27-FEB-2001; 2001US-0271855P.  
 PR 02-MAR-2001; 2001US-0272788P.  
 PR 02-MAR-2001; 2001US-0273045P.  
 PR 14-MAR-2001; 2001US-0275925P.  
 PR 14-MAR-2001; 2001US-0275947P.  
 PR 14-MAR-2001; 2001US-0275950P.  
 PR 14-MAR-2001; 2001US-0275989P.  
 PR 15-MAR-2001; 2001US-0276448P.

PR 15-MAR-2001; 2001US-0276450P.  
 PR 16-MAR-2001; 2001US-0276397P.  
 PR 16-MAR-2001; 2001US-0276768P.  
 PR 20-MAR-2001; 2001US-0278652P.  
 PR 26-MAR-2001; 2001US-0278775P.  
 PR 26-MAR-2001; 2001US-0278778P.  
 PR 29-MAR-2001; 2001US-0279882P.  
 PR 29-MAR-2001; 2001US-0279884P.  
 PR 30-MAR-2001; 2001US-0280147P.  
 PR 11-APR-2001; 2001US-0282992P.  
 PR 11-APR-2001; 2001US-0283083P.  
 PR 20-APR-2001; 2001US-0285133P.  
 PR 23-APR-2001; 2001US-0285749P.  
 PR 03-MAY-2001; 2001US-0288327P.  
 PR 03-MAY-2001; 2001US-0288504P.  
 PR 29-MAY-2001; 2001US-0294047P.  
 PR 30-MAY-2001; 2001US-0294473P.  
 PR 08-JUN-2001; 2001US-0296964P.  
 PR 18-JUN-2001; 2001US-0298959P.  
 PR 19-JUN-2001; 2001US-0299324P.  
 PR 13-AUG-2001; 2001US-0312020P.  
 PR 16-AUG-2001; 2001US-0312889P.  
 PR 16-AUG-2001; 2001US-0312908P.  
 PR 21-AUG-2001; 2001US-0313390P.  
 PR 28-AUG-2001; 2001US-0315470P.  
 PR 31-AUG-2001; 2001US-0316447P.  
 PR 07-SEP-2001; 2001US-0318115P.  
 PR 07-SEP-2001; 2001US-0318118P.  
 PR 12-SEP-2001; 2001US-0318740P.  
 PR 19-SEP-2001; 2001US-0323379P.  
 PR 18-OCT-2001; 2001US-0330245P.  
 PR 18-OCT-2001; 2001US-0330308P.  
 PR 14-NOV-2001; 2001US-0332701P.

(CURA-) CURAGEN CORP.  
 XX  
 PA Tcherenev VT, Spytek KA, Zernusen BD, Paturajan M, Shinkets RA;  
 XX Li L, Gangolli EA, Padigan M, Anderson DW, Rastelli L, Miller CE;  
 XX Gerlach V, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CE;  
 XX Futak K, Grose WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;  
 XX  
 DR WPI; 2002-706998/76.  
 XX N-PSDB; AD116507.  
 XX  
 PT New NOVX polypeptides and nucleic acids, useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 PS Claim 1; SEQ ID NO 44; 1498bp; English.

This invention relates to a novel nucleic acids, and encoded polypeptides thereof, which have properties related to the stimulation of biochemical or physiological responses in a cell, tissue, organ or organism. Specifically, it refers to the use of biologically active fragments for diagnostic and prognostic assays and furthermore in the treatment of diverse pathological conditions. The present invention describes novel human and murine NOVX proteins, as well as methods to modulate their expression using antisense oligos, ribozymes and peptide nucleic acids. The NOVX polypeptides, polynucleotides and antibodies are useful in treating or preventing NOVX-associated disorders, e.g. cardiomyopathy, atherosclerosis, cancer and diabetes. Furthermore, they may be used in treating or preventing diseases such as inflammation, autoimmune disorders, allergies, blood disorders, acquired immunodeficiency syndrome (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy and epilepsy. Accordingly, these molecules have many activities including cytosolic, cardiac, antiinflammatory, immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic, antiaesthetic, nephrotoxic, antibacterial, hepatocytotoxic, neuroprotective, nootropic, antibacterial, virucide, antiparasitic, relaxant and anticonvulsant. In addition, they are useful in screening assays to identify small molecules that modulate or inhibit, for example,

CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
 CC used as in chromosome mapping, tissue typing, preventive medicine and  
 CC pharmacogenomics. This polypeptide is a human NOVX protein of the  
 CC invention.

XX Sequence 757 AA;

Query Match 99.1%; Score 3754; DB 5; Length 757;  
 Best Local Similarity 99.3%; Pred. No. 1e-240;  
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRASIKSFVVTSTVAFPDSTKVQRTQDNCSCGFLHARGVELMRFTTPG 60  
 DB 75 MAERVVMLPPRASIKSFVVTSTVAFPDSTKVQRTQDNCSCGFLHARGVELMRFTTPG 134  
 QY 61 FPDSPYPAHARCOMALRGDADSVSLTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120  
 DB 135 FPDSPYPAHARCOMALRGDADSVSLTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 194  
 QY 121 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPMSSCGRLRKAQGTFN 180  
 DB 195 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPMSSCGRLRKAQGTFN 254  
 QY 181 SPYYPGHYPPIIDCTNMIIEVNNQHVYRFEFFYLLEPRACGTCPCDYVEINEKTCGE 240  
 DB 255 SPYYPGHYPPIIDCTNMIIEVNNQHVYRFEFFYLLEPRACGTCPCDYVEINEKTCGE 314  
 QY 241 RSQPVVTSNSNKTIVRFHSDSYTDGTFLAEYLSYDSSDPCPGQFTCRGTGCIKELRCD 300  
 DB 315 RSQPVVTSNSNKTIVRFHSDSYTDGTFLAEYLSYDSSDPCPGQFTCRGTGCIKELRCD 374  
 QY 301 GMADCTDHSDELNCSGDAHQFTCKNPKCKPLFVWCDSDVNDGSDDEQSCSPAQTFCR 360  
 DB 375 GMADCTDHSDELNCSGDAHQFTCKNPKCKPLFVWCDSDVNDGSDDEQSCSPAQTFCR 434  
 QY 361 SNGKCLSKSQOCCNGKDCGDSDEASCPKXNVVCTGTYRCLNGLCLSKGNPECDKED 420  
 DB 435 SNGKCLSKSQOCCNGKDCGDSDEASCPKXNVVCTGTYRCLNGLCLSKGNPECDKED 494  
 QY 421 CSDGDEKDCDCGSLRFTROARVVGCTDADGEMPMQVSLHALQGHICGASLISPMNLV 480  
 DB 495 CSDGDEKDCDCGSLRFTROARVVGCTDADGEMPMQVSLHALQGHICGASLISPMNLV 554  
 QY 481 SAACHYIDDRGFRTSDPTQWTAFLGLHDQSORSAPOVERLRKRIISHPEFNDFTFDYDI 540  
 DB 555 SAACHYIDDRGFRTSDPTQWTAFLGLHDQSORSAPOVERLRKRIISHPEFNDFTFDYDI 614  
 QY 541 ALLELEKPAEYSSMVRPICLPDASHVPRAGKAIWVTGHTHTQYCGTGALLIQKEIRVIN 600  
 DB 615 ALLELEKPAEYSSMVRPICLPDASHVPRAGKAIWVTGHTHTQYCGTGALLIQKEIRVIN 674  
 QY 601 QTTCENTLPOQITPRMMCVGFLSGVNSCGDSGGLPSYEADGRIFQAGVNSGDCQAQ 660  
 DB 675 QTTCENTLPOQITPRMMCVGFLSGVNSCGDSGGLPSYEADGRIFQAGVNSGDCQAQ 734  
 QY 661 RNKRGVYTRLPLFRDWIKENTGV 683  
 DB 735 RNKRGVYTRLPLFRDWIKENTGV 757

RESULT 13  
 AAY06671  
 ID AAY06671 standard; protein; 855 AA.

XX AAY06671;

DT 09-NOV-1999 (first entry)

XX Tumour antigen derived gene-15 (TADG-15) protein.

XX Tumour antigen derived gene-15; TADG-15; serine protease; human;

KM breast cancer; ovary cancer; carcinoma; diagnosis.

XX

OS Homo sapiens.

FX Key Location/Qualifiers

FT Domain 1..54

FT Domain /note= "cytoplasmic domain"

FT Domain 55..213

FT Modified-site /note= "transmembrane domain"

FT Modified-site 109..111

FT Region /note= "Aen is N-glycosylated"

FT Region 214..447

FT Region /note= "CUB repeat"

FT Modified-site 302..304

FT Region /note= "Aen is N-glycosylated"

FT Region 453..602

FT Region /note= "ligand-binding repeat (class A motif)"

FT Region 481..483

FT Region /note= "conserved SDE motif"

FT Region 518..520

FT Region /note= "conserved SDE motif"

FT Region 554..556

FT Region /note= "conserved SDE motif"

FT Region 597..599

FT Cleavage-site /note= "conserved SDE motif"

FT Domain 614..615

FT Domain /note= "catalytic domain"

FT Domain 615..855

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

FT Domain /note= "catalytic domain"

Db 293 LCGTTPPSVNLTFHSSQVNLITLITNTERHHPGEATFFOLPRMSSCGRLRAQGTFFN 352  
Qy 181 SPYYPGHYPNIDCTWNIIEVNNQHVAVRFEKPYLLBPRRACGTCPKDYVEINKEKTCGE 240  
Db 353 SPYYPGHYPNIDCTWNIIEVNNQHVAVRFEKPYLLBPRRACGTCPKDYVEINKEKTCGE 412  
Qy 241 RSQFVTVSNKTIYRFHSDSYDTGTGLAEYLSYDSDPCPGQFTCTGRCIRKELRCD 300  
Db 413 RSQFVTVSNKTIYRFHSDSYDTGTGLAEYLSYDSDPCPGQFTCTGRCIRKELRCD 472  
Qy 301 GMACTDHSDELNCSGDAHQFTCKNKECKPLFWVCDSDVNDGSDGSCSPAQTFR 360  
Db 473 GMACTDHSDELNCSGDAHQFTCKNKECKPLFWVCDSDVNDGSDGSCSPAQTFR 532  
Qy 361 SNGKCLSKSQCGNKGDCGSDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKED 420  
Db 533 SNGKCLSKSQCGNKGDCGSDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKED 592  
Qy 421 CSDGSDXDCDGLRSFTQARVVGTTADGEGWPMQVSLHALGQGHICGASLISPNWLV 480  
Db 593 CSDGSDXDCDGLRSFTQARVVGTTADGEGWPMQVSLHALGQGHICGASLISPNWLV 652  
Qy 481 SAACIYIDDRGFRYSDFQWTAFLGLHDQSORSAFGVERLKRITISHPFNDFTFDYDI 540  
Db 653 SAACIYIDDRGFRYSDFQWTAFLGLHDQSORSAFGVERLKRITISHPFNDFTFDYDI 712  
Qy 541 ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVTGNGHTQYGGTALILQKEIRVIN 600  
Db 713 ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVTGNGHTQYGGTALILQKEIRVIN 772  
Qy 601 QTTCENLIPQOITPRMNCVGLSGVDSGQSDSGPLSSVADGRIIQAQVWSMGDCAQ 660  
Db 773 QTTCENLIPQOITPRMNCVGLSGVDSGQSDSGPLSSVADGRIIQAQVWSMGDCAQ 832  
Qy 661 RNKPGVYTRPLPLFRDWIKENTGV 683  
Db 833 RNKPGVYTRPLPLFRDWIKENTGV 855

RESULT 14  
AAB98500  
ID AAB98500 standard; protein: 855 AA.  
XX AAB98500;  
AC AAB98500;  
DT 03-AUG-2001 (first entry)  
XX 03-AUG-2001 (first entry)  
DE Human TADG-15.  
XX Human; TADG-15; cytosolic; vaccine; ovarian tumour; cancer;  
KW tumour antigen-derived gene 15; extracellular serine protease.  
XX Homo sapiens.  
OS Homo sapiens.  
PN MO200129056-A1.  
PD 26-APR-2001.  
PE 20-OCT-2000; 2000MO-US029095.  
PR 20-OCT-1999; 99US-00421213.  
XX (UYAR-) UNIV ARKANSAS.  
XX O'Brien TJ, Tanimoto H;  
PI WPI; 2001-381031/40.  
DR N-PSDB; AAH23601.  
XX Novel extracellular serine protease, termed tumor antigen-derived gene 15  
PT protein overexpressed in carcinomas and DNA encoding it, for diagnosis,  
PT treatment, prevention of cancer, particularly breast, ovarian cancer.

XX Claim 11; Fig 2; 130pp; English.  
PS The present sequence represents human tumour antigen-derived gene 15.  
CC (TADG-15) protein. TADG-15 is an extracellular serine protease. It was  
CC found that TADG-15 is over-expressed in ovarian tumours. TADG-15 protein  
CC or its fragments of 9-20 residues that lack TADG-15 protease activity are  
CC useful for vaccinating an individual against TADG-15, having, suspected  
CC of having or at risk of getting cancer. Furthermore, the TADG-15 gene can  
CC be used as a diagnostic or therapeutic target in cancer  
XX  
SQ Sequence 855 AA;  
Query Match 99.1%; Score 3754; DB 4; Length 855;  
Best Local Similarity 99.3%; Pred. No. 1.2e-240;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 MAERVVMLPPRASLSFVTVSVAPPTDSKTQVQTDNCSFGLHARGVELMRFTTGG 60  
Db 173 MAERVVMLPPRASLSFVTVSVAPPTDSKTQVQTDNCSFGLHARGVELMRFTTGG 232  
Qy 61 PPDSPPYAHACQWALRGDADSVLSLTPRSFDLASCDERGSGLVTVNTLSPMEPHALVQ 120  
Db 233 PPDSPPYAHACQWALRGDADSVLSLTPRSFDLASCDERGSGLVTVNTLSPMEPHALVQ 292  
Qy 121 LCGTTPPSVNLTFHSSQVNLITLITNTERHHPGEATFFOLPRMSSCGRLRAQGTFFN 180  
Db 293 LCGTTPPSVNLTFHSSQVNLITLITNTERHHPGEATFFOLPRMSSCGRLRAQGTFFN 352  
Qy 181 SPYYPGHYPNIDCTWNIIEVNNQHVAVRFEKPYLLBPRRACGTCPKDYVEINKEKTCGE 240  
Db 353 SPYYPGHYPNIDCTWNIIEVNNQHVAVRFEKPYLLBPRRACGTCPKDYVEINKEKTCGE 412  
Qy 241 RSQFVTVSNKTIYRFHSDSYDTGTGLAEYLSYDSDPCPGQFTCTGRCIRKELRCD 300  
Db 413 RSQFVTVSNKTIYRFHSDSYDTGTGLAEYLSYDSDPCPGQFTCTGRCIRKELRCD 472  
Qy 301 GMACTDHSDELNCSGDAHQFTCKNKECKPLFWVCDSDVNDGSDGSCSPAQTFR 360  
Db 473 GMACTDHSDELNCSGDAHQFTCKNKECKPLFWVCDSDVNDGSDGSCSPAQTFR 532  
Qy 361 SNGKCLSKSQCGNKGDCGSDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKED 420  
Db 533 SNGKCLSKSQCGNKGDCGSDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKED 592  
Qy 421 CSDGSDXDCDGLRSFTQARVVGTTADGEGWPMQVSLHALGQGHICGASLISPNWLV 480  
Db 593 CSDGSDXDCDGLRSFTQARVVGTTADGEGWPMQVSLHALGQGHICGASLISPNWLV 652  
Qy 481 SAACIYIDDRGFRYSDFQWTAFLGLHDQSORSAFGVERLKRITISHPFNDFTFDYDI 540  
Db 653 SAACIYIDDRGFRYSDFQWTAFLGLHDQSORSAFGVERLKRITISHPFNDFTFDYDI 712  
Qy 541 ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVTGNGHTQYGGTALILQKEIRVIN 600  
Db 713 ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVTGNGHTQYGGTALILQKEIRVIN 772  
Qy 601 QTTCENLIPQOITPRMNCVGLSGVDSGQSDSGPLSSVADGRIIQAQVWSMGDCAQ 660  
Db 773 QTTCENLIPQOITPRMNCVGLSGVDSGQSDSGPLSSVADGRIIQAQVWSMGDCAQ 832  
Qy 661 RNKPGVYTRPLPLFRDWIKENTGV 683  
Db 833 RNKPGVYTRPLPLFRDWIKENTGV 855

RESULT 15  
AAB06930  
ID AAB06930 standard; protein: 855 AA.  
XX AAB06930;  
AC AAB06930;  
DT 16-OCT-2001 (first entry)

XX Human membrane-type serine protease (MTSP) 1.  
DE Human; transmembrane serine protease; membrane-type serine protease;  
XX MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;  
KM lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;  
XX matrilysin.  
XX Homo sapiens.  
FH Key location/Qualifiers  
FT Domain 615..855  
FT /label= Protease\_domain  
XX  
XX W0200157194-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 02-FEB-2001; 2001WO-US003471.  
XX  
XX PR 03-FEB-2000; 2000US-0179982P.  
XX PR 18-FEB-2000; 2000US-0183542P.  
XX PR 22-JUN-2000; 2000US-0213124P.  
XX PR 26-JUL-2000; 2000US-0220970P.  
XX PR 08-SEP-2000; 2000US-00657986.  
XX PR 22-SEP-2000; 2000US-0234840P.  
XX  
XX PA (CORV-) CORVAS INT INC.  
XX  
XX PI Madison EL, Ong EO, Yeh J;  
XX WPI; 2001-488877/53.  
XX DR N-PSDB; AAD13113.  
XX  
XX PT Novel single chain polypeptide comprising protease domain of type-II  
PT membrane-type serine protease or its catalytically active portion useful  
PT for treating and preventing cancer and tumor.  
XX  
XX PS Claim 12; Page 195-197; 256pp; English.  
XX  
XX CC The invention relates to transmembrane serine proteases and their  
XX corresponding nucleotides and the protease domain of a type-II membrane-  
XX type serine protease (MTSP). MTSP is useful for identifying compounds  
XX that modulate or inhibit its proteolytic activity and for formulating a  
XX medicament for treating neoplastic disease. MTSP and its corresponding  
XX nucleotides are useful in preventing or treating tumors or cancers such  
XX as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in  
XX diagnostics and in hybridisation assays. MTSP is useful as a diagnostic  
XX marker for tumour development, growth and/or progression and as  
XX immunogens to generate antibodies that specifically bind to it. MTSP DNA  
XX is useful in a yeast two-hybrid system and in gene therapy. The present  
XX sequence is human MTSP1 protein (also called matrilysin)  
SQ Sequence 855 AA;  
Query Match 99.1%; Score 3754; DB 4; Length 855;  
Best local Similarity 99.3%; Pred. No. 1.2e-240;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MAERRYVMLPPRASLSFVVTSTVVAFTDSKTYQRTQDNSCSGLIARGYELMRFTTPG 60  
DB 173 MAERRYVMLPPRASLSFVVTSTVVAFTDSKTYQRTQDNSCSGLIARGYELMRFTTPG 232  
QY 61 FPDSPYPAHARCOVALRGDADSVSLTFRSPDLASCBERGSDLVTVVNTLSPEMPEHALVQ 120  
DB 233 FPDSPYPAHARCOVALRGDADSVSLTFRSPDLASCBERGSDLVTVVNTLSPEMPEHALVQ 292  
QY 121 LCGTTPSPSYNLTFHSSQNVLLITLITWERRHPGFATFQGLPRMSSCGGLRKAQGTFN 180  
DB 293 LCGTTPSPSYNLTFHSSQNVLLITLITWERRHPGFATFQGLPRMSSCGGLRKAQGTFN 352  
QY 181 SPYRGHYPRIIDCTWNIIEVNNQHVKRFKFFLLBPRACGTCPPDYVHINEKTCGE 240  
|||||

DB 353 SPYRGHYPRIIDCTWNIIEVNNQHVKRFKFFLLBPRAGVPAGTCPPDYVEINKEKTCGE 412  
QY 241 RSQFVVTSSNKKITVRHSDQSYTDGTFLAEYLSYSSDDPCPQFTGRTGRCTRKEARCD 300  
DB 413 RSQFVVTSSNKKITVRHSDQSYTDGTFLAEYLSYSSDDPCPQFTGRTGRCTRKEARCD 472  
QY 301 GMADCTDHSDELNCSCDAGHQFTCKNFKCPFLFWVCDSVNDCCGNSDEGCSCPAQTFRFC 360  
DB 473 GMADCTDHSDELNCSCDAGHQFTCKNFKCPFLFWVCDSVNDCCGNSDEGCSCPAQTFRFC 532  
QY 361 SNGKCLSKSQOCCGKDDCGSDPASCPRKNVVTCTGHTYRCLNGCLTSGNPECDKED 420  
DB 533 SNGKCLSKSQOCCGKDDCGSDPASCPRKNVVTCTGHTYRCLNGCLTSGNPECDKED 592  
QY 421 CSDGSEKDCDCGLRSEFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480  
DB 593 CSDGSEKDCDCGLRSEFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652  
QY 481 SAHCYITDRGFYSDPTQTAFGLHDQSRSAFPGVQERRLKRIISHPFNDFTFYDI 540  
DB 653 SAHCYITDRGFYSDPTQTAFGLHDQSRSAFPGVQERRLKRIISHPFNDFTFYDI 712  
QY 541 ALLELEKPAEYSMWPRICLPDASHVFPACKAIFWVMGHTOYGGTGALILQGEIRVIN 600  
DB 713 ALLELEKPAEYSMWPRICLPDASHVFPACKAIFWVMGHTOYGGTGALILQGEIRVIN 772  
QY 601 QTTCEMLLPQOITPRMVCVFLSGVDSCCGDSGSPLSYEADGRIFQAGVSWGDCAQ 660  
DB 773 QTTCEMLLPQOITPRMVCVFLSGVDSCCGDSGSPLSYEADGRIFQAGVSWGDCAQ 832  
QY 661 RNKPGVYTRLLPLFRDWIKENTGV 683  
DB 833 RNKPGVYTRLLPLFRDWIKENTGV 855  
RESULT 16  
AAO22929  
ID AAO22929 standard; protein; 855 AA.  
XX  
XX AC AAO22929;  
XX  
XX DT 12-DEC-2002 (first entry)  
XX  
XX DE Type II transmembrane serine protease 1 protein SEQ ID No 2.  
XX  
XX KM Cytostatic; type-II membrane-type serine protease 7; MTSP7; malignancy;  
KM neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;  
KM malignant; enzyme.  
XX  
XX OS Homo sapiens.  
XX  
XX PN W0200272786-A2.  
XX  
XX PD 19-SEP-2002.  
XX  
XX PF 13-MAR-2002; 2002WO-US007903.  
XX  
XX PR 13-MAR-2001; 2001US-0275592P.  
XX  
XX PA (CORV-) CORVAS INT INC.  
XX  
XX PI Madison EL, Ong EO;  
XX  
XX DR WPI; 2002-732827/79.  
XX  
XX DR N-PSDB; AAL53444.  
XX  
XX PT New transmembrane serine protease 7 (MTSP7) polypeptide for diagnosing  
PT neoplastic diseases, monitoring tumor progress or therapeutic  
PT effectiveness, or identifying MTSP7 modulators for treating tumors or  
PT cancers.  
XX  
XX PS Disclosure; Page 172-174; 184pp; English.  
XX

CC The invention relates to a purified single or two-chain polypeptide,  
CC which comprises the protease domain of a type-II membrane-type serine  
CC protease 7 (MSP7) or its catalytically active portion. The polypeptide  
CC comprising MSP7 is useful for detecting or diagnosing a neoplastic  
CC disease, a pre-malignant lesion, a malignancy or other pathologic  
CC condition in a subject. This polypeptide is also useful for monitoring  
CC tumour (e.g. tumour of the breast, cervix, prostate, lung, ovary or  
CC colon) progress and/or therapeutic effectiveness. The inhibitor of the  
CC polypeptide or the inhibitor of the polypeptide's zymogen form is useful  
CC for treating or preventing a neoplastic disease, or tumour initiation,  
CC growth or progression, or a (pre-)malignant condition. The polypeptide or  
CC polynucleotide is also useful for identifying modulators of MSP7, which  
CC may be used to treat cancers or tumours. This sequence represents a  
CC protein of the type-II membrane-type serine protease 1 relating to the  
CC invention

XX  
XX  
SQ Sequence 855 AA;

Query Match 99.1%; Score 3754; DB 5; Length 855;

Best Local Similarity 99.3%; Pred. No. 1.2e-240; Mismatches 0; Gaps 0;

Matches 678; Conservative 0; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRATSLKSFVTVSVAPFTDSKTVQRTQDNCSPGLHARGVELMRFTTPG 60  
DB 173 MAERVVMLEPPRATSLKSFVTVSVAPFTDSKTVQRTQDNCSPGLHARGVELMRFTTPG 232  
QY 61 PPDSPPYAHARCGVALKGDADSVLSTFRSFDLASCDERGDLYTVYNTLSMPMHALVQ 120  
DB 233 PPDSPPYAHARCGVALKGDADSVLSTFRSFDLASCDERGDLYTVYNTLSMPMHALVQ 292  
QY 121 LCGYPPSYNLTFRSSQNVLLITLTITNRRHPGEATFPLPMSSCGGLRAQGFEN 180  
DB 293 LCGYPPSYNLTFRSSQNVLLITLTITNRRHPGEATFPLPMSSCGGLRAQGFEN 352  
QY 181 SPYYPGHYPPNIDCTWNIIEVPMNQHVKKVFEFYLLBERRACGTCPKDYVEINEKCYGE 240  
DB 353 SPYYPGHYPPNIDCTWNIIEVPMNQHVKKVFEFYLLBERRACGTCPKDYVEINEKCYGE 412  
QY 241 RSQGVVTSNSKIVTRFRHSQSYTDYTGFLAEYLYSDSDPCPGQFTCTGRCIRKELRCD 300  
DB 413 RSQGVVTSNSKIVTRFRHSQSYTDYTGFLAEYLYSDSDPCPGQFTCTGRCIRKELRCD 472  
QY 301 GMACTHSDRLNCSGAGHOFCTCKNFKCKPLFWVCDSVNDGNSDQSCSPAOTPRC 360  
DB 473 GMACTHSDRLNCSGAGHOFCTCKNFKCKPLFWVCDSVNDGNSDQSCSPAOTPRC 532  
QY 361 SNGKCLSKSQCGKGDGSDGSDASCPKXVVVVTCTKHTYCLNGLCLSKGNPECDGKED 420  
DB 533 SNGKCLSKSQCGKGDGSDGSDASCPKXVVVVTCTKHTYCLNGLCLSKGNPECDGKED 592  
QY 421 CSDGSDKDCDCCGLRSFTROARVVGGTDADGEMPMQVSLHALQGHICGASLISPMWLV 480  
DB 593 CSDGSDKDCDCCGLRSFTROARVVGGTDADGEMPMQVSLHALQGHICGASLISPMWLV 652  
QY 481 SAAHCYIDDGFRYSDDPTQWTAFLGLHDQSORSA PGVERLKRILISHPFNDDTPTPDI 540  
DB 653 SAAHCYIDDGFRYSDDPTQWTAFLGLHDQSORSA PGVERLKRILISHPFNDDTPTPDI 712  
QY 541 ALLELEKPAEYSSWVRPCLPDASHVPAGAKAIWYTGHTQYGTGALILQKEIRIYN 600  
DB 713 ALLELEKPAEYSSWVRPCLPDASHVPAGAKAIWYTGHTQYGTGALILQKEIRIYN 772  
QY 601 QTTCENTLLPQOITPRMNCVGLSGGVDSQCGDSGGPLSSVEADGRIFQAGVSWGDCAQ 660  
DB 773 QTTCENTLLPQOITPRMNCVGLSGGVDSQCGDSGGPLSSVEADGRIFQAGVSWGDCAQ 832  
QY 661 RNKPGVYTRLPFLFRDWIKENTGV 683  
DB 833 RNKPGVYTRLPFLFRDWIKENTGV 855

RESULT 17  
AD116816

ID AD116816 standard; protein; 855 AA.  
XX  
AC AD116816;  
XX  
DT 15-APR-2004 (first entry)  
XX  
DE Human NOVX protein homologue SegID 352.  
XX  
KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;  
KW inflammation; autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; str.  
XX  
OS Homo sapiens.  
XX  
PN W020026849-A2.  
XX  
PD 06-SEP-2002.  
XX  
PP 31-JAN-2002; 2002W0-US002785.  
XX  
PR 31-JAN-2001; 2001US-0265395P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266767P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 07-FEB-2001; 2001US-0267057P.  
PR 08-FEB-2001; 2001US-0267459P.  
PR 09-FEB-2001; 2001US-0267823P.  
PR 15-FEB-2001; 2001US-0268974P.  
PR 26-FEB-2001; 2001US-0271664P.  
PR 27-FEB-2001; 2001US-0271839P.  
PR 27-FEB-2001; 2001US-0271855P.  
PR 02-MAR-2001; 2001US-0272788P.  
PR 02-MAR-2001; 2001US-0273046P.  
PR 14-MAR-2001; 2001US-0275925P.  
PR 14-MAR-2001; 2001US-0275947P.  
PR 14-MAR-2001; 2001US-0275950P.  
PR 14-MAR-2001; 2001US-0275989P.  
PR 15-MAR-2001; 2001US-0276448P.  
PR 15-MAR-2001; 2001US-0276450P.  
PR 16-MAR-2001; 2001US-0276377P.  
PR 16-MAR-2001; 2001US-0276768P.  
PR 20-MAR-2001; 2001US-0278652P.  
PR 26-MAR-2001; 2001US-0278775P.  
PR 26-MAR-2001; 2001US-0278778P.  
PR 29-MAR-2001; 2001US-0279882P.  
PR 29-MAR-2001; 2001US-0279884P.  
PR 30-MAR-2001; 2001US-0280147P.  
PR 11-APR-2001; 2001US-0282992P.  
PR 11-APR-2001; 2001US-0283083P.  
PR 20-APR-2001; 2001US-0285133P.  
PR 23-APR-2001; 2001US-0285749P.  
PR 03-MAY-2001; 2001US-0288377P.  
PR 03-MAY-2001; 2001US-0288504P.  
PR 29-MAY-2001; 2001US-0294047P.  
PR 30-MAY-2001; 2001US-0294473P.  
PR 08-JUN-2001; 2001US-0296964P.  
PR 18-JUN-2001; 2001US-0298959P.  
PR 19-JUN-2001; 2001US-0299344P.  
PR 13-AUG-2001; 2001US-0312020P.  
PR 16-AUG-2001; 2001US-0312889P.  
PR 16-AUG-2001; 2001US-0312908P.  
PR 21-AUG-2001; 2001US-0313390P.  
PR 28-AUG-2001; 2001US-0315470P.  
PR 31-AUG-2001; 2001US-0316447P.  
PR 07-SEP-2001; 2001US-0318115P.  
PR 07-SEP-2001; 2001US-0318118P.  
PR 12-SEP-2001; 2001US-0318740P.  
PR 19-SEP-2001; 2001US-0323379P.



PR 15-MAR-2001; 2001US-0276448P.  
 PR 15-MAR-2001; 2001US-0276450P.  
 PR 16-MAR-2001; 2001US-0276397P.  
 PR 16-MAR-2001; 2001US-0276768P.  
 PR 20-MAR-2001; 2001US-0278652P.  
 PR 26-MAR-2001; 2001US-0278775P.  
 PR 26-MAR-2001; 2001US-0278778P.  
 PR 29-MAR-2001; 2001US-0279882P.  
 PR 29-MAR-2001; 2001US-0279884P.  
 PR 30-MAR-2001; 2001US-0280147P.  
 PR 11-APR-2001; 2001US-0282992P.  
 PR 11-APR-2001; 2001US-0283083P.  
 PR 20-APR-2001; 2001US-0285133P.  
 PR 23-APR-2001; 2001US-0285749P.  
 PR 03-MAY-2001; 2001US-0288327P.  
 PR 03-MAY-2001; 2001US-0288504P.  
 PR 29-MAY-2001; 2001US-0294047P.  
 PR 30-MAY-2001; 2001US-0294473P.  
 PR 08-JUN-2001; 2001US-0296964P.  
 PR 18-JUN-2001; 2001US-0298959P.  
 PR 19-JUN-2001; 2001US-0299324P.  
 PR 13-AUG-2001; 2001US-0312020P.  
 PR 16-AUG-2001; 2001US-0312889P.  
 PR 16-AUG-2001; 2001US-0312908P.  
 PR 21-AUG-2001; 2001US-0313390P.  
 PR 28-AUG-2001; 2001US-0315470P.  
 PR 31-AUG-2001; 2001US-0316447P.  
 PR 07-SEP-2001; 2001US-0318115P.  
 PR 07-SEP-2001; 2001US-0318118P.  
 PR 12-SEP-2001; 2001US-0318740P.  
 PR 19-SEP-2001; 2001US-0323379P.  
 PR 18-OCT-2001; 2001US-0330245P.  
 PR 18-OCT-2001; 2001US-0330308P.  
 PR 14-NOV-2001; 2001US-0332701P.

XX (CURA-) CURAGEN CORP.

PA Tchernev VT, Spytek KA, Zethusen BD, Paturajan M, Shinketa RA;  
 PI Li L, Gangoli EA, Padigaru M, Anderson DM, Rastelli L, Miller CE;  
 PI Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CEA,  
 PI Futak K, Grosse WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;  
 XX WPI; 2002-706998/76.

XX New NOVX polypeptides and nucleic acids, useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.

PS Disclosure; SEQ ID NO 418; 1498bp; English.

XX This invention relates to a novel nucleic acid, and encoded polypeptides  
 CC thereof, which have properties related to the stimulation of biochemical  
 CC or physiological responses in a cell, tissue, organ or organism.  
 CC Specifically, it refers to the use of biologically active fragments for  
 CC diagnostic and prognostic assays and furthermore in the treatment of  
 CC diverse pathological conditions. The present invention describes novel  
 CC human and murine NOVX proteins, as well as methods to modulate their  
 CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
 CC The NOVX polypeptides, polynucleotides and antibodies are useful in  
 CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,  
 CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
 CC treating or preventing diseases such as inflammation, autoimmune  
 CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
 CC (AIDS), obesity, asthma, immunoglobulin (Ig) A nephropathy, cirrhosis,  
 CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy  
 CC and epilepsy. Accordingly, these molecules have many activities including  
 CC cytoskeletal, cardiac, antiinflammatory, immunosuppressive, antiallergic,  
 CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,  
 CC antiasthmatic, nephroprotective, antiarrhythmic, hepatocytic,  
 CC neuroprotective, neurotropic, antibacterial, vincine, antiparasitic,  
 CC relaxant and anticonvulsant. In addition, they are useful in screening  
 CC assays to identify small molecules that modulate or inhibit, for example,

CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
 CC used as in chromosome mapping, tissue typing, preventive medicine and  
 CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein  
 CC of the invention.

XX Sequence 855 AA;

Query Match 99.1%; Score 3754; DB 5; Length 855;  
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;  
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRARSLKSPVVTSVVAPEPTDSTVQRTQDNCSCFGLHARGVELMRFTTPG 60  
 DB 173 MAERVVMLEPPRARSLKSPVVTSVVAPEPTDSTVQRTQDNCSCFGLHARGVELMRFTTPG 232  
 QY 61 PPDSPPYPAHRCQMALRGDADSVLSLTFRSPDLASCDBERSDLYVTYNTLSPMEPALVQ 120  
 DB 233 PPDSPPYPAHRCQMALRGDADSVLSLTFRSPDLASCDBERSDLYVTYNTLSPMEPALVQ 292  
 QY 121 LCGTYPSPYNTLTPHSSONVLLITLTNTERRHGFEPFQLPRMSSCGRLRAQGTFN 180  
 DB 293 LCGTYPSPYNTLTPHSSONVLLITLTNTERRHGFEPFQLPRMSSCGRLRAQGTFN 352  
 QY 181 SPYYPGHYPNIDCTWNIENVNQHVRKFFYLLEPRRACGTCPKDYVEINGEKYCGE 240  
 DB 353 SPYYPGHYPNIDCTWNIENVNQHVRKFFYLLEPRRACGTCPKDYVEINGEKYCGE 412  
 QY 241 RSQFVVTNSNKTIVRHSQSTYTDGFLAEYLSYDSDDCPQFTQRTGRCIRKELRCD 300  
 DB 413 RSQFVVTNSNKTIVRHSQSTYTDGFLAEYLSYDSDDCPQFTQRTGRCIRKELRCD 472  
 QY 301 GMADCTHSDBLNCSCDAGHOFCKNFKCKPLFWVCDSVNDCCGNSDEGSCSPAQTFRC 360  
 DB 473 GMADCTHSDBLNCSCDAGHOFCKNFKCKPLFWVCDSVNDCCGNSDEGSCSPAQTFRC 532  
 QY 361 SNGKCLSKSQOQNGKDDCGDSDEASCPRXNVVYCTKHTYRCNLGLCLSKGNPECDKED 420  
 DB 533 SNGKCLSKSQOQNGKDDCGDSDEASCPRXNVVYCTKHTYRCNLGLCLSKGNPECDKED 592  
 QY 421 CSDGSDKDDCCGTRSTRTRQARVVGCTDADBEQPMQVSLHALGGHICGASLISPMWLV 480  
 DB 593 CSDGSDKDDCCGTRSTRTRQARVVGCTDADBEQPMQVSLHALGGHICGASLISPMWLV 652  
 QY 481 SAACHYIDDRGFRSPDPTQMTAFGLHDOSORSAPOQBERLRKRITSHPPFNFTFPYDI 540  
 DB 653 SAACHYIDDRGFRSPDPTQMTAFGLHDOSORSAPOQBERLRKRITSHPPFNFTFPYDI 712  
 QY 541 ALLELEKPAEYSSNVRPICLPDASHVFPACKAIWVTGHTQYGGTGALLILOGEIRVIN 600  
 DB 713 ALLELEKPAEYSSNVRPICLPDASHVFPACKAIWVTGHTQYGGTGALLILOGEIRVIN 772  
 QY 601 QTTCENTLPPQITTPRMKCVFSLSGVDSCGDSGSPISVEADGRIFQAGVSWGDCAQ 660  
 DB 773 QTTCENTLPPQITTPRMKCVFSLSGVDSCGDSGSPISVEADGRIFQAGVSWGDCAQ 832  
 QY 661 RNKPGVYTRLPFPDWTKENTGV 683  
 DB 833 RNKPGVYTRLPFPDWTKENTGV 855

# RESULT 19

ID AD116875 standard; protein: 855 AA.

XX AD116875;

XX 15-Apr-2004 (first entry)

XX Human NOVX protein homologue SeqID 411.

XX human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;

KW inflammation; autoimmune disorder; allergy; blood disorder;

KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;



KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
KM Alzheimer's disease; infection; str.  
XX Homo sapiens.  
XX  
PN WO200268649-A2.  
XX  
PD 06-SEP-2002.  
XX  
PF 31-JAN-2002; 2002WO-US002785.  
XX  
PR 31-JAN-2001; 2001US-0265395P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266767P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 07-FEB-2001; 2001US-0267057P.  
PR 08-FEB-2001; 2001US-0267459P.  
PR 09-FEB-2001; 2001US-0267823P.  
PR 15-FEB-2001; 2001US-0268974P.  
PR 26-FEB-2001; 2001US-0271664P.  
PR 27-FEB-2001; 2001US-0271839P.  
PR 27-FEB-2001; 2001US-0271855P.  
PR 02-MAR-2001; 2001US-0272788P.  
PR 02-MAR-2001; 2001US-0273046P.  
PR 14-MAR-2001; 2001US-0275925P.  
PR 14-MAR-2001; 2001US-0275947P.  
PR 14-MAR-2001; 2001US-0275950P.  
PR 14-MAR-2001; 2001US-0275989P.  
PR 15-MAR-2001; 2001US-0276448P.  
PR 15-MAR-2001; 2001US-0276450P.  
PR 16-MAR-2001; 2001US-0276397P.  
PR 16-MAR-2001; 2001US-0276768P.  
PR 20-MAR-2001; 2001US-0278652P.  
PR 26-MAR-2001; 2001US-0287752P.  
PR 26-MAR-2001; 2001US-0287778P.  
PR 29-MAR-2001; 2001US-0279882P.  
PR 29-MAR-2001; 2001US-0279884P.  
PR 30-MAR-2001; 2001US-0280147P.  
PR 11-APR-2001; 2001US-0282982P.  
PR 11-APR-2001; 2001US-0283083P.  
PR 20-APR-2001; 2001US-0285133P.  
PR 23-APR-2001; 2001US-0285749P.  
PR 03-MAY-2001; 2001US-0288327P.  
PR 03-MAY-2001; 2001US-0288504P.  
PR 29-MAY-2001; 2001US-0294047P.  
PR 30-MAY-2001; 2001US-0294473P.  
PR 08-JUN-2001; 2001US-0296964P.  
PR 18-JUN-2001; 2001US-0298959P.  
PR 19-JUN-2001; 2001US-0299324P.  
PR 13-AUG-2001; 2001US-0312020P.  
PR 16-AUG-2001; 2001US-0312889P.  
PR 16-AUG-2001; 2001US-0312908P.  
PR 21-AUG-2001; 2001US-0313390P.  
PR 28-AUG-2001; 2001US-0315470P.  
PR 31-AUG-2001; 2001US-0316447P.  
PR 07-SEP-2001; 2001US-0318115P.  
PR 07-SEP-2001; 2001US-0318118P.  
PR 12-SEP-2001; 2001US-0318740P.  
PR 19-SEP-2001; 2001US-0323379P.  
PR 18-OCT-2001; 2001US-0330245P.  
PR 18-OCT-2001; 2001US-0330308P.  
PR 14-NOV-2001; 2001US-0332701P.  
XX  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Tchernev VT, Spletke KA, Zethusen BD, Paturajan M, Shinkets RA;  
PI Li L, Ganggoli EA, Padigaru M, Anderson DW, Rascelli L, Miller CE;  
PI Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CRA;  
PI Pirtak K, Grose WM, Alsbrook JF, Lepley DM, Rieger DK, Burgess CE;  
XX

DR WPI, 2002-706998/76.  
XX  
XX New NOVX polypeptides and nucleic acids, useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX  
XX Disclosure; SEQ ID NO 411; 1498bp; English.  
XX  
XX This invention relates to a novel nucleic acids, and encoded polypeptides  
CC thereof, which have properties related to the stimulation of biochemical  
CC or physiological responses in a cell, tissue, organ or organism.  
CC Specifically, it refers to the use of biologically active fragments for  
CC diagnostic and prognostic assays and furthermore in the treatment of  
CC diverse pathological conditions. The present invention describes novel  
CC human and murine NOVX proteins, as well as methods to modulate their  
CC expression using antisense oligos, ribozymes and peptide nucleic acids.  
CC The NOVX polypeptides, polynucleotides and antibodies are useful in  
CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,  
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in  
CC treating or preventing diseases such as inflammation, autoimmune  
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome  
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,  
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy,  
CC and epilepsy. Accordingly, these molecules have many activities including  
CC cytostatic, cardiac, antiinflammatory, immunosuppressive, antiallergic,  
CC hemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,  
CC antisthmatic, neurotropic, antichratic, hepatotropic,  
CC neuroprotective, nootropic, antibacterial, viricide, antiparasitic,  
CC relaxant and anticonvulsant. In addition, they are useful in screening  
CC assays to identify small molecules that modulate or inhibit, for example,  
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also  
CC used as in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein  
CC of the invention.  
XX  
XX Sequence 855 AA;  
SQ  
Query Match 99.1%; Score 3754; DB 5; Length 855;  
Best Local Similarity 99.3%; Pred. No. 1,2e-240;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MAERVYMLPPRRSLKSVYTSVVAFPPTDSKVQRTQDNSCFGHARGVEMRTTTC 60  
DB 173 MAERVYMLPPRRSLKSVYTSVVAFPPTDSKVQRTQDNSCFGHARGVEMRTTTC 232  
QY 61 FPDSPYPAHRCOMALRGDASVLSLTPRSFDLASCDEKSDLVYTYNTLSPMERPAIVQ 120  
DB 233 FPDSPYPAHRCOMALRGDASVLSLTPRSFDLASCDEKSDLVYTYNTLSPMERPAIVQ 292  
QY 121 LCGTYPSPYNLTFHSSQVLLITLITNTERRHGFEATFPQJLRMSCCGRLKKAQGTEN 180  
DB 293 LCGTYPSPYNLTFHSSQVLLITLITNTERRHGFEATFPQJLRMSCCGRLKKAQGTEN 352  
QY 181 SPYYPGHYPPNIDCTNIEVPNNQHKVAFKFFYLLEPPRACCTCKDVEINIGEXYCE 240  
DB 353 SPYYPGHYPPNIDCTNIEVPNNQHKVAFKFFYLLEPPRACCTCKDVEINIGEXYCE 412  
QY 241 RSQFVVTNSNKTITVFHSDQSYTDGFLAEVLSYSSPPCQFPCRTGRCTRKLRPCD 300  
DB 413 RSQFVVTNSNKTITVFHSDQSYTDGFLAEVLSYSSPPCQFPCRTGRCTRKLRPCD 472  
QY 301 GWADCTDHSDELNCSGDAGHPTCKNKFCKPLFWVCDSVNDCGNSDEGCGCPAQTFFC 360  
DB 473 GWADCTDHSDELNCSGDAGHPTCKNKFCKPLFWVCDSVNDCGNSDEGCGCPAQTFFC 532  
QY 361 SNGKCLSKSQCCNGKDCDGDSPDASCPRKNNVYTCRHTYRCLNGCLSKNPECCGKXD 420  
DB 533 SNGKCLSKSQCCNGKDCDGDSPDASCPRKNNVYTCRHTYRCLNGCLSKNPECCGKXD 592  
QY 421 CSQSDKXKDCDGLRFTROARVVGSTDADEGEMPVQVSLHALGQSHIGASLISPNMTV 480  
DB 593 CSQSDKXKDCDGLRFTROARVVGSTDADEGEMPVQVSLHALGQSHIGASLISPNMTV 652

QY 481 SAACHYIDDRGFRYSDDPTQWTAFLGLHDQSORSAQVQERLKRILISHPFNDFTPDYDI 540  
 DB 653 SAACHYIDDRGFRYSDDPTQWTAFLGLHDQSORSAQVQERLKRILISHPFNDFTPDYDI 712  
 QY 541 ALLELEKPAEYSSWVRPICLPDASHVPFAGKAIWVTGHTQYGGTGLILQKEIRVIN 600  
 DB 713 ALLELEKPAEYSSWVRPICLPDASHVPFAGKAIWVTGHTQYGGTGLILQKEIRVIN 772  
 QY 601 QTTCENLIPQOITPRMNCVGFSLGGVDS CGDGGPLSSVEADGRIFQAGVSWGDCQAQ 660  
 DB 773 QTTCENLIPQOITPRMNCVGFSLGGVDS CGDGGPLSSVEADGRIFQAGVSWGDCQAQ 832  
 QY 661 RNKPGVYTRLPFRDWMKENTGV 683  
 DB 833 RNKPGVYTRLPFRDWMKENTGV 855

RESULT 20  
 ABP56619  
 ID ABP56619 standard; proteoin; 855 AA.  
 XX  
 AC ABP56619;  
 DT 24-MAR-2003 (first entry)  
 XX  
 DE Human membrane-type serine protease MTSPI protein SEQ ID NO:2.  
 XX  
 KM Human; membrane-type serine protease; enzyme; MTSPI0; cytosolic;  
 KM type-II membrane-type serine protease; neoplastic disease; tumour; MTSPI;  
 KM matrixase.  
 XX  
 OS Homo sapiens.  
 PN WO200292841-A2.  
 PD 21-NOV-2002.  
 PF 14-MAY-2002; 2002W0-US015332.  
 PR 14-MAY-2001; 2001US-0291001P.  
 PA (CORV-) CORVAS INT INC.  
 PI Madison BL, Yeh J;  
 DR WPI; 2003-129309/12.  
 DR N-PSDB; ABZ22450.  
 XX  
 PT New polypeptides comprising the protease domain of a type-II membrane-  
 PT type serine protease (MTSPI0), or its mutants, useful for diagnosing  
 PT neoplasms or malignancies, or for screening for MTSPI0 inhibitors for  
 PT treating such diseases.  
 XX  
 PS Disclosure; Page 181-183; 198pp; English.  
 XX  
 CC The present invention describes a polypeptide comprising a purified  
 CC single or two chain polypeptide, which comprises the protease domain of a  
 CC type-II membrane-type serine protease (MTSPI0) or its catalytically  
 CC active portion, or a mutant of it, where up to 50 % of the amino acids  
 CC are replaced with another amino acid, and the resulting polypeptide is a  
 CC single chain or two chain polypeptide that has a catalytic activity of at  
 CC least 1-10 % of the unmutated polypeptide. MTSPI0 has cytosolic  
 CC activity. The polypeptide containing the protease domain of the MTSPI0 is  
 CC useful for detecting a neoplastic disease, and for diagnosing the  
 CC presence of a pre-malignant lesion, a malignancy, or other pathologic  
 CC condition in a subject, or monitoring tumour (e.g. breast, cervix,  
 CC prostate, lung, ovary or colon tumour) progression and/or therapeutic  
 CC effectiveness. An inhibitor of the polypeptide containing the protease  
 CC domain of MTSPI0 is useful for treating or preventing neoplastic disease  
 CC in a mammal. An inhibitor of the activation cleavage of the zymogen form  
 CC of the MTSPI0 polypeptide is useful for inhibiting tumour initiation,  
 CC growth or progression, or treating (pre-)malignant conditions of the e.g.

CC breast, cervix, prostate, lung, ovary or colon. The present sequence  
 CC represents human MTSPI (also known as matrixase), which is used in an  
 CC example from the present invention  
 XX  
 SQ Sequence 855 AA;  
 Query Match 99.1%; Score 3754; DB 6; Length 855;  
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;  
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 MAERVYMLPPRARSLKSPVTSVAPPTDSKTVOQTODNSCSFGILHARGVELMRFTTPG 60  
 DB 173 MAERVYMLPPRARSLKSPVTSVAPPTDSKTVOQTODNSCSFGILHARGVELMRFTTPG 232  
 QY 61 PPDSPYPAHARCCQALRGDADSVLSLTFRRSFDLASCBERSDLYVTYNTLSPMEPALVQ 120  
 DB 233 PPDSPYPAHARCCQALRGDADSVLSLTFRRSFDLASCBERSDLYVTYNTLSPMEPALVQ 292  
 QY 121 LCGTYPPSYNLTFHSSQNVLLITLTITWTERRHGFETAFQPLRMSCGRLRKAQGTFN 180  
 DB 293 LCGTYPPSYNLTFHSSQNVLLITLTITWTERRHGFETAFQPLRMSCGRLRKAQGTFN 352  
 QY 181 SPYYPGHYPNNIDCTWNIIEVNNQHVVRKFFYLLEPRRACGTCPRDYVEINGEKYCGE 240  
 DB 353 SPYYPGHYPNNIDCTWNIIEVNNQHVVRKFFYLLEPRRACGTCPRDYVEINGEKYCGE 412  
 QY 241 RSQFVVTSSNKTITVRHSQSYTDGTGFLAEVLSYDSDPCPQFTQRTGRCIRKELRCD 300  
 DB 413 RSQFVVTSSNKTITVRHSQSYTDGTGFLAEVLSYDSDPCPQFTQRTGRCIRKELRCD 472  
 QY 301 GMADCTHSDDELNCS CDAGHOFCKNFKCPFLFVVCDSVNDGDNSEDGSCSPAQTFRC 360  
 DB 473 GMADCTHSDDELNCS CDAGHOFCKNFKCPFLFVVCDSVNDGDNSEDGSCSPAQTFRC 532  
 QY 361 SNGKCLSKSQCKNGKDCGDSDEASCPRKXNVVYCTKHYRCCLNGCLSKGNBECDEKED 420  
 DB 533 SNGKCLSKSQCKNGKDCGDSDEASCPRKXNVVYCTKHYRCCLNGCLSKGNBECDEKED 592  
 QY 421 CSQSDSEKDCDCGRSTTRQARVVGCTDADGEMPMQVSLHALGCGHICGASLISPMWLV 480  
 DB 593 CSQSDSEKDCDCGRSTTRQARVVGCTDADGEMPMQVSLHALGCGHICGASLISPMWLV 652  
 QY 481 SAACHYIDDRGFRYSDDPTQWTAFLGLHDQSORSAQVQERLKRILISHPFNDFTPDYDI 540  
 DB 653 SAACHYIDDRGFRYSDDPTQWTAFLGLHDQSORSAQVQERLKRILISHPFNDFTPDYDI 712  
 QY 541 ALLELEKPAEYSSWVRPICLPDASHVPFAGKAIWVTGHTQYGGTGLILQKEIRVIN 600  
 DB 713 ALLELEKPAEYSSWVRPICLPDASHVPFAGKAIWVTGHTQYGGTGLILQKEIRVIN 772  
 QY 601 QTTCENLIPQOITPRMNCVGFSLGGVDS CGDGGPLSSVEADGRIFQAGVSWGDCQAQ 660  
 DB 773 QTTCENLIPQOITPRMNCVGFSLGGVDS CGDGGPLSSVEADGRIFQAGVSWGDCQAQ 832  
 QY 661 RNKPGVYTRLPFRDWMKENTGV 683  
 DB 833 RNKPGVYTRLPFRDWMKENTGV 855

RESULT 21  
 ID AAO30146  
 AAO30146 standard; proteoin; 855 AA.  
 XX  
 AC AAO30146;  
 DT 03-SEP-2003 (first entry)  
 XX  
 DE Human membrane-type serine protease MTSPI protein.  
 XX  
 KM Serine protease 17; CYSPI7; tumour; cancer; antiense therapy; prostate;  
 KM breast; cervix; lung; ovary; colon; gene therapy; human; enzyme; MTSPI;  
 KM membrane-type serine protease; matrixase.

OS Homo sapiens.  
 XX  
 PN WO2003044179-A2.  
 PD 30-MAY-2003.  
 XX PF 20-NOV-2002; 2002WO-US037626.  
 XX PR 20-NOV-2001; 2001US-0332015P.  
 XX PA (CORV-) CORVAS INT. INC.  
 XX PI Madison EL, Ong EO;  
 XX DR WPI; 2003-449816/42.  
 DR N-PsDB; AAL60792.  
 XX  
 PT New substantially purified serine protease 17 polypeptide and encoding  
 PT nucleic acid, useful for diagnosing and treating tumor conditions and/or  
 PT cancer, particularly of the breast, cervix, prostate, lung, ovary or  
 PT colon.  
 XX  
 PS Disclosure; Page 175-177; 1899p; English.  
 XX  
 CC The invention relates to serine protease 17 polypeptide designated CVSP17  
 CC and its corresponding nucleic acid sequence. The invention also relates  
 CC to a method using CVSP17 protein to identify compounds that modulate its  
 CC protease activity. The method is useful for preventing, diagnosing and  
 CC treating disorders related to the serine protease 17 activity, such as  
 CC tumour conditions and/or cancer, particularly of the breast, prostate,  
 CC cervix, lung, ovary or colon. CVSP17 DNA is used in gene therapy and in  
 CC antisense therapy. The present sequence is human membrane-type serine  
 CC protease MTSPI (also called matrilysin) protein. This sequence is used to  
 CC illustrate the method of the invention  
 XX  
 SQ Sequence 855 AA;  
 Query Match 99.1%; Score 3754; DB 6; Length 855;  
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;  
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 MAERVVMLPPRRSLKSFVVTSSVAFPTDSKYVORTQDNSCSFGLHARGVEIMRFTTPG 60  
 DB 173 MAERVVMLPPRRSLKSFVVTSSVAFPTDSKYVORTQDNSCSFGLHARGVEIMRFTTPG 232  
 QY 61 FPDSPYPAHARCGWALRGDADSVLTFRSPDLASCPERGSDDLVTYNTLSPMHPALVQ 120  
 DB 233 FPDSPYPAHARCGWALRGDADSVLTFRSPDLASCPERGSDDLVTYNTLSPMHPALVQ 292  
 QY 121 LCGTTPSYNLTFFHSQNVLLITLITNTERRHPPGFATFPOLPRMSSCGRLRKAQGTFFN 180  
 DB 293 LCGTTPSYNLTFFHSQNVLLITLITNTERRHPPGFATFPOLPRMSSCGRLRKAQGTFFN 352  
 QY 181 SPYYPGHPYPRNIDCTWNIIEVPNNOHVVRKFFFLLEPRRACGTCPODYVEINKEKTCGE 240  
 DB 353 SPYYPGHPYPRNIDCTWNIIEVPNNOHVVRKFFFLLEPRRACGTCPODYVEINKEKTCGE 412  
 QY 241 RSPGVVTSNKNKTVTRHSPQSYVDGFLAEVLSYDSSDPCPGFQTRTKRCKTRKELRCD 300  
 DB 413 RSPGVVTSNKNKTVTRHSPQSYVDGFLAEVLSYDSSDPCPGFQTRTKRCKTRKELRCD 472  
 QY 301 GWADCTDHSDELNCSGAGHGFCTCKNFKCKPLFVWCDVNDGCGNSDEGSCSPAQTFRFC 360  
 DB 473 GWADCTDHSDELNCSGAGHGFCTCKNFKCKPLFVWCDVNDGCGNSDEGSCSPAQTFRFC 532  
 QY 361 SNGKCLSKSQOQCNKGKDCGGSDEASCPKVVVVTCTHTTRCLNGLCLSKGNPECKED 420  
 DB 533 SNGKCLSKSQOQCNKGKDCGGSDEASCPKVVVVTCTHTTRCLNGLCLSKGNPECKED 592  
 QY 421 CSPGSDKDDCDGCRFTROARVVGFTDADGEMPMWVSHALGGGHTCGASLISPMWLV 480  
 DB 593 CSPGSDKDDCDGCRFTROARVVGFTDADGEMPMWVSHALGGGHTCGASLISPMWLV 652

QY 481 SAACHYIDRGRFSDPTQWTAFLGLHDQSQRSAPOVERRLKRIISHPEFNDFTEDYDI 540  
 DB 653 SAACHYIDRGRFSDPTQWTAFLGLHDQSQRSAPOVERRLKRIISHPEFNDFTEDYDI 712  
 QY 541 ALLELEKPAEYSSMWPRICLPDASHVFPACKAIWWTGMGHTQYGGTGALLILOGEIRVIN 600  
 DB 713 ALLELEKPAEYSSMWPRICLPDASHVFPACKAIWWTGMGHTQYGGTGALLILOGEIRVIN 772  
 QY 601 QTTCEMLLPQOITPRMWCVFLSGVDSGCGDSGCPSSVEADGRIFOAGVSMGDCAQ 660  
 DB 773 QTTCEMLLPQOITPRMWCVFLSGVDSGCGDSGCPSSVEADGRIFOAGVSMGDCAQ 832  
 QY 661 RNKPGYTRLPFLFRDWIKENTGV 683  
 DB 833 RNKPGYTRLPFLFRDWIKENTGV 855  
 RESULT 22  
 AAE29820  
 ID AAE29820 standard; protein; 855 AA.  
 XX  
 AC AAE29820;  
 XX  
 DT 24-FEB-2003 (first entry)  
 XX  
 DE Human membrane-type serine protease 1 (MTSP1).  
 XX  
 KW Human; type II membrane-type serine protease 9; tumour; transgenic;  
 KW type II transmembrane serine protease; enzyme; gene therapy; MTSP9;  
 KW neoplastic disease; transgenic animal; membrane-type serine protease 1;  
 KW TSP; MTSP1; matrilysin.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200277267-A2.  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-US009611.  
 XX  
 PR 27-MAR-2001; 2001US-0279228P.  
 PR 15-MAY-2001; 2001US-0291501P.  
 XX  
 PA (CORV-) CORVAS INT. INC.  
 XX  
 PI Madison EL, Ong EO;  
 XX  
 DR WPI; 2003-018940/01.  
 DR N-PsDB; AAD47225.  
 XX  
 PT New substantially purified single or two-chain type II membrane-type  
 PT serine protease 9 (MTSP9) polypeptide, useful for monitoring tumor  
 PT progression, inhibiting tumor initiation, or treating a malignant or pre-  
 PT malignant condition.  
 XX  
 PS Disclosure; Page 183-185; 199p; English.  
 XX  
 CC The invention relates to type II membrane-type serine protease 9 (MTSP9)  
 CC polypeptides and polymucleotides. MTSP belongs to type II transmembrane  
 CC serine protease (TSP) family. Sequences of the invention and their  
 CC antibodies are useful for diagnosing, treating or preventing neoplastic  
 CC disease in mammals. They are useful for monitoring tumor progression,  
 CC inhibiting tumor initiation, growth or progression or treating malignant  
 CC or pre-malignant conditions. Transgenic animals of the invention are  
 CC useful in animal models of tumor initiation, growth and/or progression  
 CC models. The invention is also useful in gene therapy. The present  
 CC sequence is human membrane-type serine protease 1 (MTSP1). MTSP1 also  
 CC referred as matrilysin is a member of the TSP family  
 XX  
 SQ Sequence 855 AA;  
 Query Match 99.1%; Score 3754; DB 6; Length 855;  
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;

Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRARSLSFVVTSVAFPTDSKTVQRTODNSCSFGHARVGLMRFTTPG 60  
DB 173 MAERVVMLPPRARSLSFVVTSVAFPTDSKTVQRTODNSCSFGHARVGLMRFTTPG 232  
QY 61 FPDSPYPAHARCQWALRGDADSVSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVQ 120  
DB 233 FPDSPYPAHARCQWALRGDADSVSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVQ 292  
QY 121 LCGTTPPSYNLTFHSSQNVLLITLITNTERRHGFEPATFQLPMSSCGRLRAQGTEN 180  
DB 293 LCGTTPPSYNLTFHSSQNVLLITLITNTERRHGFEPATFQLPMSSCGRLRAQGTEN 352  
QY 181 SPYYPGHYPNIDCTNMIIEVNNQHVYKRFKFLLEBPRAACGCPDYVEINEKTCGE 240  
DB 353 SPYYPGHYPNIDCTNMIIEVNNQHVYKRFKFLLEBPRAACGCPDYVEINEKTCGE 412  
QY 241 RSQFVTSNSNKTIVRFHSDSYTDGFLAEYLSYSSDPCPGQFTCRTRCIRKELRCD 300  
DB 413 RSQFVTSNSNKTIVRFHSDSYTDGFLAEYLSYSSDPCPGQFTCRTRCIRKELRCD 472  
QY 301 GMADCTDHSDELNCSCDAGHQTCKNFKCPLEFWVCDVNDGDNDSDEQSCCPAQTFRC 360  
DB 473 GMADCTDHSDELNCSCDAGHQTCKNFKCPLEFWVCDVNDGDNDSDEQSCCPAQTFRC 532  
QY 361 SNGKCLSKSQOCCNGKDCGDSDEASCPRKXNVVTCIKTRKCLNGLCLSKNPECCKED 420  
DB 533 SNGKCLSKSQOCCNGKDCGDSDEASCPRKXNVVTCIKTRKCLNGLCLSKNPECCKED 592  
QY 421 CSDGSDKDCDCGRLSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLSPMWLV 480  
DB 593 CSDGSDKDCDCGRLSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLSPMWLV 652  
QY 481 SAACHYIDDRGFRYSPTQWTAFLGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDI 540  
DB 653 SAACHYIDDRGFRYSPTQWTAFLGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDI 712  
QY 541 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWVGWGHYQYGGTALLIQKEIRVIN 600  
DB 713 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWVGWGHYQYGGTALLIQKEIRVIN 772  
QY 601 QTTCENTLLPOQITPRMMCVGFLSGVDSCOGDSGPISSVEADGRIFQAGVSWGDCQAQ 660  
DB 773 QTTCENTLLPOQITPRMMCVGFLSGVDSCOGDSGPISSVEADGRIFQAGVSWGDCQAQ 832  
QY 661 RNKPGVYTRLPFLFRDWIKENTGV 683  
DB 833 RNKPGVYTRLPFLFRDWIKENTGV 855

RESULT 23  
AAE29791  
ID AAE29791 standard; protein: 855 AA.  
XX  
AC AAE29791:  
XX  
XX 24-FEB-2003 (first entry)  
XX  
DE Human membrane-type serine protease, MTSPL.  
XX  
KM Human; serine protease 14; CYSPI4; cancer; malignancy; breast; colon;  
XX gene therapy; cytosolic; membrane-type serine protease; enzyme; MTSPL.  
XX  
OS Homo sapiens.  
XX  
PN WO200277263-A2.  
XX  
PD 03-OCT-2002.  
XX  
XX 20-MAR-2002; 2002WO-US009039.  
XX  
XX 22-MAR-2001; 2001US-0278166P.  
PR

XX  
PA (CORV-) CORVAS INT INC.  
XX  
PI Madison EL, Yeh J;  
XX  
XX WPI: 2003-018938/01.  
DR N-PSDB; AAD47180.  
XX  
XX  
PT New purified CYSPI4 polypeptide and encoding nucleic acid molecule,  
PT useful for diagnosing, preventing and/or treating disorders, such as  
PT cancers and malignancies of the breast, cervix, prostate, lung, ovary or  
PT colon.  
XX  
PS Disclosure; Page 171-173; 185pp; English.

XX  
XX The invention relates to transmembrane serine protease 14 (CYSPI4), its  
XX nucleic acid sequence and the method based on them. The methods and  
XX compositions of the invention are useful for diagnosing, preventing  
XX and/or treating conditions associated with the aberrant expression or  
XX activity of the CYSPI4 polypeptide, such as cancers and malignancies of  
XX the breast, cervix, prostate, lung, ovary or colon. The methods are also  
XX useful for identifying compounds that will modulate the protease activity  
XX of CYSPI4 polypeptide, and monitoring tumor progression and/or  
XX therapeutic effectiveness. CYSPI4 DNA used in gene therapy. The present  
XX sequence is human membrane-type serine protease, MTSPL

SQ Sequence 855 AA.

Query Match 99.1%; Score 3754; DB 6; Length 855;  
Best Local Similarity 99.3%; Pred. No. 1.2e-240;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRARSLSFVVTSVAFPTDSKTVQRTODNSCSFGHARVGLMRFTTPG 60  
DB 173 MAERVVMLPPRARSLSFVVTSVAFPTDSKTVQRTODNSCSFGHARVGLMRFTTPG 232  
QY 61 FPDSPYPAHARCQWALRGDADSVSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVQ 120  
DB 233 FPDSPYPAHARCQWALRGDADSVSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVQ 292  
QY 121 LCGTTPPSYNLTFHSSQNVLLITLITNTERRHGFEPATFQLPMSSCGRLRAQGTEN 180  
DB 293 LCGTTPPSYNLTFHSSQNVLLITLITNTERRHGFEPATFQLPMSSCGRLRAQGTEN 352  
QY 181 SPYYPGHYPNIDCTNMIIEVNNQHVYKRFKFLLEBPRAACGCPDYVEINEKTCGE 240  
DB 353 SPYYPGHYPNIDCTNMIIEVNNQHVYKRFKFLLEBPRAACGCPDYVEINEKTCGE 412  
QY 241 RSQFVTSNSNKTIVRFHSDSYTDGFLAEYLSYSSDPCPGQFTCRTRCIRKELRCD 300  
DB 413 RSQFVTSNSNKTIVRFHSDSYTDGFLAEYLSYSSDPCPGQFTCRTRCIRKELRCD 472  
QY 301 GMADCTDHSDELNCSCDAGHQTCKNFKCPLEFWVCDVNDGDNDSDEQSCCPAQTFRC 360  
DB 473 GMADCTDHSDELNCSCDAGHQTCKNFKCPLEFWVCDVNDGDNDSDEQSCCPAQTFRC 532  
QY 421 CSDGSDKDCDCGRLSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLSPMWLV 480  
DB 593 CSDGSDKDCDCGRLSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLSPMWLV 652  
QY 481 SAACHYIDDRGFRYSPTQWTAFLGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDI 540  
DB 653 SAACHYIDDRGFRYSPTQWTAFLGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDI 712  
QY 541 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWVGWGHYQYGGTALLIQKEIRVIN 600  
DB 713 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWVGWGHYQYGGTALLIQKEIRVIN 772  
QY 601 QTTCENTLLPOQITPRMMCVGFLSGVDSCOGDSGPISSVEADGRIFQAGVSWGDCQAQ 660

Db 773 OTTCENLIPQOITPRMNCVGLSGVDS CGDGGFLSVEADGRIFQAGVSWMGCAQ 832  
 QY 661 RNRPGVYTRLPFRDWIKENTGV 683  
 Db 833 RNRPGVYTRLPFRDWIKENTGV 855

## RESULT 24

ABP72376  
 ID ABP72376 standard; protein; 855 AA.

AC ABP72376;

DT 13-MAY-2003 (first entry)

DE Transmembrane serine protease 1 (MSP1).

XX Human; transmembrane serine protease 1; MSP1; MSP20; enzyme;

KW cytostatic; dermatological; cardiant; vulnerary; ophthalmological;

KM gene therapy.

OS Homo sapiens.

PN WO2003004681-A2.

PD 16-JAN-2003.

PF 03-JUL-2002; 2002WO-US021208.

PR 03-JUL-2001; 2001US-0302939P.

PA (CORV-) CORVAS INT INC.

PI Madison EL, Ong EO;

XX WPI; 2003-239207/23.

DR N-PADB; ABZ58500.

XX New type-II membrane-type serine protease 20 polypeptides, useful for

PT preparing a medicament for diagnosing, treating or preventing cancer,

PT dermatological disorders, aberrant wound repairs or crest syndromes.

XX Disclosure; Page 198-200; 216pp; English.

XX The present sequence is the protein sequence of human type II

CC transmembrane serine protease 1 (MSP1). The invention relates to novel

CC human type II transmembrane serine protein 20 (MSP20) (see ABP72374) and

CC nucleic acids encoding it (see ABZ58499). Also claimed are methods of

CC inhibiting tumour initiation, growth or progression by inhibiting MSP20

CC activity, and of treating or preventing a disease or disorder associated

CC with undesired and/or uncontrolled angiogenesis or neovascularisation,

CC especially undesired angiogenesis associated with solid neoplasms,

CC vascular malformations and cardiovascular disorders, chronic inflammatory

CC diseases, aberrant wound repairs, circulatory disorders, crest syndromes,

CC dermatological disorders and ocular disorders using an inhibitor of MSP20

XX Sequence 855 AA;

XX

XX Query Match 99.1%; Score 3754; DB 6; Length 855;

XX Best Local Similarity 99.3%; Pred. No. 1.2e-240;

XX Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRARSLSFVVTSVAPPTSKYQRTQDNSSCFGLHARGVELMRFTTPG 60

Db 173 MAERVVMLPPRARSLSFVVTSVAPPTSKYQRTQDNSSCFGLHARGVELMRFTTPG 232

QY 61 FPDSPYPAHARCGWALRGDADSVLSLTFRSFDLASCDERSDLYTVYNTLSPMEPHALVQ 120

Db 233 FPDSPYPAHARCGWALRGDADSVLSLTFRSFDLASCDERSDLYTVYNTLSPMEPHALVQ 292

QY 121 LCGTTPPSVNLTFHSSQNVLLITLINTERRHGFEATFFQLPRMSSCGRLKKAQGTEN 180

Db 293 LCGTTPPSVNLTFHSSQNVLLITLINTERRHGFEATFFQLPRMSSCGRLKKAQGTEN 352  
 QY 181 SPYYPGHYPNIDCTNIEVPNNQHVAFKFFYLLEPRRACGTCPODYVEINGEKXCGE 240  
 Db 353 SPYYPGHYPNIDCTNIEVPNNQHVAFKFFYLLEPRRACGTCPODYVEINGEKXCGE 412  
 QY 241 RSGFVVTSSNKNITVAFHSDQSYDTGFLAEYLSDSDPCPOQFTCRTRCIRKELRCD 300  
 Db 413 RSGFVVTSSNKNITVAFHSDQSYDTGFLAEYLSDSDPCPOQFTCRTRCIRKELRCD 472  
 QY 301 GMADCTDHSDELNCSGDAGHOFCKNRFCKPLFWVCDSVNDGDNDEGCGCPAOTFRC 360  
 Db 473 GMADCTDHSDELNCSGDAGHOFCKNRFCKPLFWVCDSVNDGDNDEGCGCPAOTFRC 532  
 QY 361 SNGKCLSKSQCGKNGKDCDGSDEASCPRKYNVYTCFKHYRCINGLSKGNPECCKED 420  
 Db 533 SNGKCLSKSQCGKNGKDCDGSDEASCPRKYNVYTCFKHYRCINGLSKGNPECCKED 592  
 QY 421 CSDGDEKDCDGLRSFTROARVVGCTDADGEGMPQVSLHALGQHCILSGASLISPMILV 480  
 Db 593 CSDGDEKDCDGLRSFTROARVVGCTDADGEGMPQVSLHALGQHCILSGASLISPMILV 652  
 QY 481 SAHICYIDRGRRYSPTQWTAFLGLHDQSQSPAPGVORRLKRIISHPFENDFTFDYI 540  
 Db 653 SAHICYIDRGRRYSPTQWTAFLGLHDQSQSPAPGVORRLKRIISHPFENDFTFDYI 712  
 QY 541 ALLELEKPAEYSSMVPICLPDASHVFPAGKAIWYTGWHTQYGTGALILQGEIRVIN 600  
 Db 713 ALLELEKPAEYSSMVPICLPDASHVFPAGKAIWYTGWHTQYGTGALILQGEIRVIN 772  
 QY 601 OTTCENLIPQOITPRMNCVGLSGVDS CGDGGFLSVEADGRIFQAGVSWMGCAQ 660  
 Db 773 OTTCENLIPQOITPRMNCVGLSGVDS CGDGGFLSVEADGRIFQAGVSWMGCAQ 832  
 QY 661 RNRPGVYTRLPFRDWIKENTGV 683  
 Db 833 RNRPGVYTRLPFRDWIKENTGV 855

## RESULT 25

ADB97551  
 ID ADB97551 standard; protein; 855 AA.

XX ADB97551;

AC 04-DEC-2003 (first entry)

DT Human MSP1, SEQ ID NO:2.

XX Human; type I transmembrane serine protease 1; MSP1; matrilysin;

KW serine protease; protease cleavage activation; diagnostic marker;

KM neoplastic disorder; cancer; breast; cervix; prostate; lung; ovary;

KW cytostatic; gene therapy; drug screening; tumour progression; monitoring;

XX enzyme.

XX Homo sapiens.

XX WO2003031585-A2.

XX 17-APR-2003.

XX 08-OCT-2002; 2002WO-US032417.

XX 09-OCT-2001; 2001US-0328530P.

XX (CORV-) CORVAS INT INC.

XX Madison EL, Yeh J;

XX WPI; 2003-393442/37.

XX N-PADB; ADB97550.

XX New purified single- or two-chain polypeptide, useful for diagnosing,

PT preventing or treating cancer (e.g. colon cancer), comprises a protease  
 PT domain of a type-1 membrane-type serine protease 25 or its catalytically  
 PT active portion.  
 PS Disclosure; Page 179-181; 97pp; English.  
 XX  
 CC The invention relates to human type I transmembrane serine protease 25  
 CC (MSP25; AD897565) and polypeptides derived from it (e.g., AD897555). The  
 CC MSP25 gene is located on chromosome 12. MSP25 is a serine protease  
 CC which is activated by protease cleavage to yield a two-chain protease  
 CC comprising an A chain and a B chain linked by a disulphide bond. MSP25  
 CC is expressed or is active in tumour cells, and can therefore be used as a  
 CC diagnostic marker for certain cancers. The invention also encompasses  
 CC nucleic acids encoding an MSP25 polypeptide (AD897564, AD897571).  
 CC AD897554; nucleic acid vectors and host cells comprising an MSP25  
 CC polynucleotide; a MSP25 knockout animal; and an antibody specific for  
 CC either the single chain (zymogen) or two-chain (activated) form of  
 CC MSP25. MSP25 polypeptides are useful in diagnosing, preventing or  
 CC treating neoplastic diseases, such as cancer of the breast, cervix,  
 CC prostate, lung, ovary or colon. MSP25 polypeptides may also be used for  
 CC identifying compounds that modulate the protease activity of the  
 CC polypeptide and for monitoring tumour progression and/or therapeutic  
 CC effectiveness. The present sequence represents the related protein, MTSPI  
 CC (also known as matrilysin).  
 CC  
 XX  
 SQ Sequence 855 AA;

Query Match 99.1%; Score 3754; DB 7; Length 855;  
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;  
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRARSLSKSFVTVSVAPPTDSKTVOPTDONSCEFGHARGVELMRFTTPG 60  
 DB 173 MAERVVMLEPPRARSLSKSFVTVSVAPPTDSKTVOPTDONSCEFGHARGVELMRFTTPG 232  
 QY 61 FPDSPYPAHARCMALRGDADSVLTFERSFDLASCEBERGSDLVTVNTLSPMEPHALVQ 120  
 DB 233 FPDSPYPAHARCMALRGDADSVLTFERSFDLASCEBERGSDLVTVNTLSPMEPHALVQ 292  
 QY 121 LCGTYPSPYNTLFHSSQNVLLITLITNTERRHPEFATFQGLPMSSCGGLRRAQGTFN 180  
 DB 293 LCGTYPSPYNTLFHSSQNVLLITLITNTERRHPEFATFQGLPMSSCGGLRRAQGTFN 352  
 QY 181 SPYRGHYPPNIDCTNMTIEVNNQNVVRKFEFLASPRACGTCPPDYVEINEKTCGE 240  
 DB 353 SPYRGHYPPNIDCTNMTIEVNNQNVVRKFEFLASPRACGTCPPDYVEINEKTCGE 412  
 QY 241 RSQFVTVSNENKLTFRFHSDDSYDTGFLAEVLSYDSSDPCPGQFTCRGTGCIKEIRCD 300  
 DB 413 RSQFVTVSNENKLTFRFHSDDSYDTGFLAEVLSYDSSDPCPGQFTCRGTGCIKEIRCD 472  
 QY 301 GMADCTHSDDELNSCDAGHQFTCKNPKCKPLFWVCDVNDGDNDEGSCCPAQTFRC 360  
 DB 473 GMADCTHSDDELNSCDAGHQFTCKNPKCKPLFWVCDVNDGDNDEGSCCPAQTFRC 532  
 QY 361 SNGKCLSKSQOQCNCKDCCGSDASCPKVVVCTGTRCNGLCLSKGNPCDCKED 420  
 DB 533 SNGKCLSKSQOQCNCKDCCGSDASCPKVVVCTGTRCNGLCLSKGNPCDCKED 592  
 QY 421 CSDSDSDKDCDCCGARSFTQARVVGGTDADGEMPMOVSJHALGQGHICGASLSIPMWLV 480  
 DB 593 CSDSDSDKDCDCCGARSFTQARVVGGTDADGEMPMOVSJHALGQGHICGASLSIPMWLV 652  
 QY 481 SAACHYIDDRGFYSPTQTWATFGLHDQGRSAPGVQERLRKIIISHPFNDFTFDYDI 540  
 DB 653 SAACHYIDDRGFYSPTQTWATFGLHDQGRSAPGVQERLRKIIISHPFNDFTFDYDI 712  
 QY 541 ALLELEKPARYSNVPRICLPDASHVPRAGKAIWVTGHTQYCGTALTIQKEIRIVIN 600  
 DB 713 ALLELEKPARYSNVPRICLPDASHVPRAGKAIWVTGHTQYCGTALTIQKEIRIVIN 772  
 QY 601 QTTCENTLLPOQITPRMNCVGFSLGGVNSCGDSDGAPLSVEADGRIFQAGVNSGDDCAQ 660

DB 773 QTTCENTLLPOQITPRMNCVGFSLGGVNSCGDSDGAPLSVEADGRIFQAGVNSGDDCAQ 832  
 QY 661 RNKPGVYTRLPFRDWTAKENTGV 683  
 DB 833 RNKPGVYTRLPFRDWTAKENTGV 855

## RESULT 26

AD110371  
 ID AD110371 standard; protein; 855 AA.

XX  
 AC AD110371;

XX  
 DT 22-Apr-2004 (first entry)

XX  
 DE Human cell surface protease #1.

XX  
 KW therapeutic agent; plasmin; protease specific antigen; PSA;

XX  
 KW cell-surface protease-associated disease; cancer; ocular disease;

XX  
 KW cardiovascular disease; chronic inflammatory disease; wound;

XX  
 KW circulatory disorder; dermatological disorder; rheumatoid arthritis;

XX  
 KW psoriasis; diabetic retinopathy; pterygium;

XX  
 KW excimer laser surgery scarring; glaucoma filtering surgery scarring;

XX  
 KW melanoma; Kaposi's sarcoma; human; cell surface protease.

XX  
 OS Homo sapiens.

XX  
 XX WO200295007-A2.

XX  
 XX 23-MAY-2002; 2002NO-USO16819.

XX  
 XX 23-MAY-2001; 2001US-0293267P.

XX  
 XX (CORV-) CORVAS INT INC.

XX  
 XX Madison EL, Sempke JE, Vlaeuk GP, Kemp SJ, Komandla M, Siev DV;

XX  
 XX WPI; 2003-221280/21.

XX  
 XX DR N-PSDB; AD110370.

XX  
 XX Claim 9; SEQ ID NO 2; 581pp; English.

XX  
 CC The invention comprises a conjugate that consists of a therapeutic agent

XX  
 CC and a peptide substrate (optionally linked via linker). The peptide

XX  
 CC substrate is proteolytically cleaved by a cell surface protease pr a

XX  
 CC soluble, released or shed form of it, to liberate the therapeutic agent,

XX  
 CC the conjugate of the invention is not substantially cleaved by plasmin or

XX  
 CC protease specific antigen (PSA). The conjugate of the invention is useful

XX  
 CC for treating a cell-surface protease-associated disease such as: cancer,

XX  
 CC ocular diseases, cardiovascular diseases, chronic inflammatory diseases,

XX  
 CC wounds, circulatory disorders, dermatological disorders, rheumatoid

XX  
 CC arthritis, psoriasis, diabetic retinopathies, recurrence of pterygium,

XX  
 CC scarring from excimer laser surgery, scarring from glaucoma filtering

XX  
 CC surgery, macular degeneration, cress syndromes, solid neoplasms, vascular

XX  
 CC tumours, melanoma and Kaposi's sarcoma. The present amino acid sequence

XX  
 XX represents a human cell surface protease.

XX  
 SQ Sequence 855 AA;

Query Match 99.1%; Score 3754; DB 7; Length 855;  
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;  
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRARSLSKSFVTVSVAPPTDSKTVOPTDONSCEFGHARGVELMRFTTPG 60  
 DB 173 MAERVVMLEPPRARSLSKSFVTVSVAPPTDSKTVOPTDONSCEFGHARGVELMRFTTPG 232

Oy	61	PPDSYPAPARCOMALRGADSVLSLTFFSPDLASCDDEGSDLVITYVNTLSPEMHALVQ	120
Dd	233	FPDSYPAPARCOMALRGADSVLSLTFFSPDLASCDERSDLYVNTLSPEMHALVQ	232
Oy	121	LCGTYPSPYNLTFFHSSQNVLLITLLINTERRHHPGFATEFQDLPRMSSCGRLKKAQTFN	180
Dd	293	LCGTYPSPYNLTFFHSSQNVLLITLLINTERRHHPGFATEFQDLPRMSSCGRLKKAQTFN	352
Oy	161	SPYYGHHYEPNIDCTNIEVPNNQHYKAPKFFYLLEPPRAAGTCPCPDYIEINGEKYCGE	240
Dd	353	SPYYGHHYEPNIDCTNIEVPNNQHYKAPKFFYLLEPPRAAGTCPCPDYIEINGEKYCGE	412
Oy	241	RSQFVYVTSNKLITVYFHSDOSYTDGPFLAEYLSYDSSDPCCPGQFCTRGCIRKELRCD	300
Dd	413	RSQFVYVTSNKLITVYFHSDOSYTDGPFLAEYLSYDSSDPCCPGQFCTRGCIRKELRCD	472
Oy	301	GNADCTDHSDELNCSGDAGHQFTCKNKFCKPLFWCDVNDCCGNSDEQCSCPAQTFRC	360
Dd	473	GNADCTDHSDELNCSGDAGHQFTCKNKFCKPLFWCDVNDCCGNSDEQCSCPAQTFRC	532
Oy	361	SNGKCLSKSQCCNGKDCGPDGSDDEASCPRKVNVTCTKHTRYCLNGCLSKGNPECCKED	420
Dd	533	SNGKCLSKSQCCNGKDCGPDGSDDEASCPRKVNVTCTKHTRYCLNGCLSKGNPECCKED	592
Oy	421	CGDGSDEKDCDGLRSFTFRQARVVGTTDADDEGEMPMQVSLHALGGCHIIGASLISPMWLV	480
Dd	593	CGDGSDEKDCDGLRSFTFRQARVVGTTDADDEGEMPMQVSLHALGGCHIIGASLISPMWLV	652
Oy	481	SAHACIYIDRGRFRYSDDPTQWTAFLGLHDSQSRSAFGVQERRLKRIISHPEFNDFTFDYDI	540
Dd	653	SAHACIYIDRGRFRYSDDPTQWTAFLGLHDSQSRSAFGVQERRLKRIISHPEFNDFTFDYDI	712
Oy	541	ALLELEKPAEYSSMWRPCLCPDASHVFPAPAKAIWVTGMGHTTQYGGTGALLILOKGEIRVIN	600
Dd	713	ALLELEKPAEYSSMWRPCLCPDASHVFPAPAKAIWVTGMGHTTQYGGTGALLILOKGEIRVIN	772
Oy	601	OTTCCENLPLPQQLTTPRMKCVGPLSGGVDSGCGSDGSPFLSSVEADGRIFQAGVSMGSDCAQ	660
Dd	773	OTTCCENLPLPQQLTTPRMKCVGPLSGGVDSGCGSDGSPFLSSVEADGRIFQAGVSMGSDCAQ	832
Oy	661	RNKPQVYTRLPLFRDWIKENTGV	683
Dd	833	RNKPQVYTRLPLFRDWIKENTGV	855
RESULT 27			
ADG65326			
ADG65326 standard; protein; 855 AA.			
XX	AC	ADG65326;	
XX	AC		
XX	DE	11-MAR-2004 (first entry)	
XX	DE		
XX	XX	Human MTSP1.	
XX	XX		
KW	KW	human; type II membrane-type serine protease 12; MTSPl2; chromosome 19;	
KW	KW	protease domain; MTSPl2-PD1, MTSPl2-PD2, MTSPl2-PD3; lung; oesophageal;	
KW	KW	prostate; colon; ovary; cervix; breast; pancreas; cancer; tumour;	
XX	XX	neoplastic condition.	
OS	OS	Homo sapiens.	
XX	XX		
PH	PH	Key	Location/Qualifiers
FT	FT	Domain	615..855
XX	XX		/label = Protease domain
XX	XX	WO2003104394-A2.	
PD	PD	18-DEC-2003.	
XX	XX		
PF	PF	21-MAY-2003; 2003WO-US016181.	
XX	XX		

PR		21-MAY-2002; 2002US-0382851P.
XX		(DEND-) DENDBRON SAN DIEGO LLC.
PA		
XX		
Pt	Madison EL, Ong EO;	
XX		
DR	WPI, 2004-062325/06.	
XX		
PT	New type II membrane-type serine protease 12 proteins and nucleic acids,	
PT	useful in diagnostics, particularly for diagnosing lung, esophageal,	
PT	prostate, colon, ovary, cervix, breast and pancreas cancers.	
XX		
PS	Disclosure; SEQ ID NO 2; 207bp; English.	
XX		
CC	This sequence represents a human type II membrane-type serine protease 1	
CC	(MTSP1). The MTSP1 protein sequence was used in the isolation of the	
CC	coding sequence of MTSP12, which has been isolated to chromosome 19.	
CC	MTSP12 does not include the sequence of amino acids Arg-Lys-His-Leu-Pro-	
CC	Arg-Pro-Ala ADG65347. The MTSP12 coding sequence was identified by using	
CC	the protein sequence of the protease domain of MTSP1 to search the human	
CC	HTGS database which produced three serine proteases, MTSP12-PDI, MTSP12-	
CC	PD2 and MTSP12-PD3. Esrr's were identified which matched portions of	
CC	MTSP12-PDI, MTSP12-PD2 and MTSP12-PD3. MTSP12 polypeptides, proteins and	
CC	nucleic acids are useful in diagnostics, particularly for diagnosing	
CC	lung, oesophageal, prostate, colon, ovary, cervix, breast and pancreas	
CC	cancers. These are useful in immunoassays to detect, prognosis, diagnose,	
CC	or monitor various conditions, diseases, and disorders affecting MTSP12	
CC	polypeptide expression, or monitor their treatment. Modulators of MTSP12	
CC	are useful for treating cancer, tumour and other neoplastic conditions.	
XX		
SQ	Sequence 855 AA;	
	Query Match 99.1%; Score 3754; DB 8; Length 855;	
	Best Local Similarity 99.3%; Pred. No. 1.2e-240;	
	Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	
OY	1 MAERRVWLPPRRASLKSFFVTYSVAAPTDKTYQRTODNCSRLHARGYLEMFTTPG 60	
Db	173 MAERRVWLPPRRASLKSFFVTYSVAAPTDKTYQRTODNCSRLHARGYLEMFTTPG 232	
OY	61 PPDSPPYPAHRCQALRGDADSVSLTFRSEFDLASCDERSGDLTVVVYNTLSPMEBHALVQ 120	
Db	233 PPDSPPYPAHRCQALRGDADSVSLTFRSEFDLASCDERSGDLTVVVYNTLSPMEBHALVQ 292	
OY	121 LCGTYPBSYNLTFFHSNQVLLITLITNTERRHPGEAFATFOLPRMS CGGRLRKAOGTFN 180	
Db	293 LCGTYPBSYNLTFFHSNQVLLITLITNTERRHPGEAFATFOLPRMS CGGRLRKAOGTFN 352	
OY	181 SPYYPGHYPPINDCTWNLFEVNNOHVKKRFKPFYLLEBRBAOCNCPKYVELNGEKYGE 240	
Db	353 SPYYPGHYPPINDCTWNLFEVNNOHVKKRFKPFYLLEBRBAOCNCPKYVELNGEKYGE 412	
OY	241 RSQGFVIVSNSNKIIVRFPHSDSYTDGTGLAEYLSYSSDPGQFTCRGTCIRKEALCD 300	
Db	413 RSQGFVIVSNSNKIIVRFPHSDSYTDGTGLAEYLSYSSDPGQFTCRGTCIRKEALCD 472	
OY	301 GMACCTTHSDIDLNSCDAHQFTCKNRCKPLFWVCDSVNDCCGNSDDQGSCPAQTORC 360	
Db	473 GMACCTTHSDIDLNSCDAHQFTCKNRCKPLFWVCDSVNDCCGNSDDQGSCPAQTORC 532	
OY	361 SNGKCLSKSQOCCNGKDCCGDSDBASCPKWVWVCTKTATYCLNGLCLSKKNPECDGKED 420	
Db	533 SNGKCLSKSQOCCNGKDCCGDSDBASCPKWVWVCTKTATYCLNGLCLSKKNPECDGKED 592	
OY	421 CSDSDEKDCDGLRSPTROARVVGGTDADGEMPMOVSLHALCOGHI CGASLISPNMLV 480	
Db	593 CSDSDEKDCDGLRSPTROARVVGGTDADGEMPMOVSLHALCOGHI CGASLISPNMLV 652	
OY	481 SAACHYIDDRFRYSDDPTQWTAFLGLHDQOSRSA PGVOERTLKRI IISHPFNDPFEDVI 540	
Db	653 SAACHYIDDRFRYSDDPTQWTAFLGLHDQOSRSA PGVOERTLKRI IISHPFNDPFEDVI 712	
OY	541 ALALELEKPAEYSSWVRPICLPDASHVPAGCAI WWTGWHTQYGATGALL OKGEIRVIN 600	



```

Db      713 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGNGHTQYCGTALLIQKEIRVIN 772
Qy      601 QTTCENLTPQOITPRMNCVGFSLSGVDS CGDSCGSPSSVEADGRITFOAGVWSMGDCQAQ 660
Db      773 QTTCENLTPQOITPRMNCVGFSLSGVDS CGDSCGSPSSVEADGRITFOAGVWSMGDCQAQ 832
Qy      661 RNKGVYTRLP LFRDWIKENTGV 683
Db      833 RNKGVYTRLP LFRDWIKENTGV 855

RESULT 28
ADJ28861
ID ADJ28861 standard; protein; 855 AA.
XX
AC ADJ28861;
XX
DT 22-Apr-2004 (first entry)
XX
DE Human matricinase ( MTSP1) serine protease.
XX
KM Human; matricinase; MTSP1; CUSP16; cytolethic; gene therapy; tumour;
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 615..855
   /note="Protease domain"
XX
PN W02004005471-A2.
XX
PD 15-JAN-2004.
XX
PF 01-JUL-2003; 2003WO-US020959.
XX
PR 02-JUL-2002; 2002US-0394347P.
XX
PA (DEND-) DENDREON SAN DIEGO LLC.
XX
PI Madison EL, Ong EO, Yeh J;
XX
DR WPI; 2004-099379/10.
   N-PSDB; ADJ28860.
XX
PT New serine protease 16 polypeptides, useful for diagnosing, preventing or
   treating cancer (e.g. breast cancer) or for identifying compounds that
   may be used for modulating the polypeptides.
XX
PS Disclosure; SEQ ID NO 2; 205bp; English.
XX
CC The present sequence is the protein sequence of human MTSP1 (matricinase),
   a serine protease that is expressed in epithelial cancer and normal
   tissue. In an example from the invention, the protein sequence of the
   protease domain of MTSP1 was used to search a human genome database. A
   protease was identified and designated serine protease 16 or CUSP16.
   CUSP16 cDNA ADJ28863 was subsequently cloned from a human liver cDNA
   library by PCR. CUSP16 polypeptides exhibit protease activity as a single
   chain or as a multi-chain form. Methods are provided for identifying
   compounds that modulate the protease activity. CUSP16 polypeptides also
   serve as tumour markers.
XX
SQ Sequence 855 AA;
XX
Query Match 99.1%; Score 3754; DB 8; Length 855;
Best Local Similarity 99.3%; Pred. No. 1.2e-240;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy      1 MAERVVMLPFRASLSKSVVTSVVAFPDTSKTVQRTQDNCSCFGLHARGVELMRFTTPG 60
Db      173 MAERVVMLPFRASLSKSVVTSVVAFPDTSKTVQRTQDNCSCFGLHARGVELMRFTTPG 232

```

```

Qy      61 PPSPPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMPEHALVQ 120
Db      233 PPSPPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMPEHALVQ 292
Qy      121 LCGTTPPSYNLTFRHSQNVLLITLITNTERRHQFEATFFQLRMSGGRRLRAQGTEN 180
Db      293 LCGTTPPSYNLTFRHSQNVLLITLITNTERRHQFEATFFQLRMSGGRRLRAQGTEN 352
Qy      181 SPYTPGHYPPNIDCTWNIIEVPNNQHVVRKFPYLLPRAAGCTCPDPYIEINGEKYCGE 240
Db      353 SPYTPGHYPPNIDCTWNIIEVPNNQHVVRKFPYLLPRAAGCTCPDPYIEINGEKYCGE 412
Qy      241 RSQFVVTNSNKTITVRHSDQSYTDPTGFLAEVLSYDSDDCPQGFRTGRICRKELRCD 300
Db      413 RSQFVVTNSNKTITVRHSDQSYTDPTGFLAEVLSYDSDDCPQGFRTGRICRKELRCD 472
Qy      301 GMADCTDHSDELNCSCDAGHQFTCKNFKCLPFWCDSVND CGDSDSCSCSPAQTFRC 360
Db      473 GMADCTDHSDELNCSCDAGHQFTCKNFKCLPFWCDSVND CGDSDSCSCSPAQTFRC 532
Qy      361 SNGKCLSKSQCCNGKDCGGSDBASCPRKNVYTCTKHTYRCINGLCLSKGNECDCKED 420
Db      533 SNGKCLSKSQCCNGKDCGGSDBASCPRKNVYTCTKHTYRCINGLCLSKGNECDCKED 592
Qy      421 CSQSDSEKDCDCGLRSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLSPMWLV 480
Db      593 CSQSDSEKDCDCGLRSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLSPMWLV 652
Qy      481 SAACHYIDRGRFYSDEPTQWTAFLGLHDQSRSAPGVQERRLKRIISHPFNDFTFYDI 540
Db      653 SAACHYIDRGRFYSDEPTQWTAFLGLHDQSRSAPGVQERRLKRIISHPFNDFTFYDI 712
Qy      541 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGNGHTQYCGTALLIQKEIRVIN 600
Db      713 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGNGHTQYCGTALLIQKEIRVIN 772
Qy      601 QTTCENLTPQOITPRMNCVGFSLSGVDS CGDSCGSPSSVEADGRITFOAGVWSMGDCQAQ 660
Db      773 QTTCENLTPQOITPRMNCVGFSLSGVDS CGDSCGSPSSVEADGRITFOAGVWSMGDCQAQ 832
Qy      661 RNKGVYTRLP LFRDWIKENTGV 683
Db      833 RNKGVYTRLP LFRDWIKENTGV 855

RESULT 29
ADJ46895
ID ADJ46895 standard; protein; 855 AA.
XX
AC ADJ46895;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human transmembrane serine protease (MTSP) polypeptide #1.
XX
CC Human; transmembrane serine protease; MTSP; cell surface protease;
   plasmun; prostate specific antigen; PSA; proliferative disease;
   cell-surface protease-associated disease; autoimmune disease;
   inflammatory disease; infectious disease; endocrine disease; cancer;
   ocular disorder; cardiovascular disorder; chronic inflammatory disease;
   wound; circulatory disorder; dermatological disorder; restenosis;
   rheumatoid arthritis; psoriasis; diabetic retinopathy;
   laser surgery scarring; glaucoma filtering surgery scarring;
   macular degeneration; CRST syndrome; bacterial infection; viral disease;
   solid neoplasm; vascular tumour; lung; colon; prostate; melanoma;
   Kaposi's sarcoma; enzyme.
XX
OS Homo sapiens.
XX
PN US2004001801-A1.
XX
PD 01-JAN-2004.
XX

```

PF 23-MAY-2002; 2002US-00156214.  
XX  
PR 23-MAY-2002; 2002US-00156214.  
XX  
PA (CORV-) CORVAS INT INC.  
XX  
PI Madison EL, Semple JE, Vlaeuk GP, Kemp SJ, Komandla M, Slev DV,  
XX WPI; 2004-190126/18.  
DR N-PSDB; ADJ46894.  
XX  
PT Conjugate useful for treating e.g. cancer, cell-surface protease-  
PT associated diseases, comprising a peptidic substrate or nucleic acid  
PT substrate linked to a therapeutic agent through a linker.  
PS  
PS Claim 9; SEQ ID NO 2; 361pp; English.  
XX  
XX The invention relates to a conjugate comprising a therapeutic agent and a  
CC peptidic substrate or nucleic acid substrate linked to the agent  
CC optionally through a linker or peptidic linker, where the peptidic  
CC substrate is proteolytically cleaved by a cell surface protease or a  
CC soluble, released or shed form conjugate to liberate the agent and the  
CC conjugate is not substantially cleaved by plasmin or prostate specific  
CC antigen (PSA). The conjugate is useful for treating a disease, which  
CC involves administering a conjugate to a subject, where the disease is  
CC preferably a proliferative disease or a cell-surface protease-associated  
CC disease. The diseases include autoimmune diseases, inflammatory diseases,  
CC infectious diseases and endocrine diseases. The conjugate is useful for  
CC treating a cell-surface protease-associated disease, which involves  
CC administering a conjugate comprising an agent and a peptidic substrate to  
CC a subject exhibiting symptoms of a cell-surface protease-associated  
CC disorder, where the disease is selected from cancer, ocular disorders,  
CC cardiovascular disorders, chronic inflammatory diseases, wounds,  
CC circulatory disorders, dermatological disorders, restenosis, rheumatoid  
CC arthritis, psoriasis, diabetic retinopathies, scarring from laser  
CC surgery, scarring from glaucoma filtering surgery, macular degeneration,  
CC CRIST syndrome, bacterial infections, viral diseases, solid neoplasms and  
CC melanoma and Kaposi's sarcoma. This sequence represents a transmembrane  
CC serine protease (MTSP) polypeptide of the invention.  
XX  
SQ Sequence 855 AA;  
Query Match 99.1%; Score 3754; DB 8; Length 855;  
Best Local Similarity 99.3%; Pred. No. 1.2e-240;  
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MAERVVMLEPPRARSLKSFVVTSVAPPTSTKYQRTQDSCSGFLHARGVELMRFTTGG 60  
DB 173 MAERVVMLEPPRARSLKSFVVTSVAPPTSTKYQRTQDSCSGFLHARGVELMRFTTGG 232  
QY 61 FPDSPYPAHARCGWALRGDASVLSLTFRSFDLASCDERSGDLVTVNTLSPEPHALVQ 120  
DB 233 FPDSPYPAHARCGWALRGDASVLSLTFRSFDLASCDERSGDLVTVNTLSPEPHALVQ 292  
QY 121 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHQFEATFQLPMSSCGRLKRAQGTEN 180  
DB 293 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHQFEATFQLPMSSCGRLKRAQGTEN 352  
QY 181 SPVYPGHYPPNIDCTNWIIEVNNQHVVRKFFYLEPRACGTCPPDYVEINKEKCGE 240  
DB 353 SPVYPGHYPPNIDCTNWIIEVNNQHVVRKFFYLEPRACGTCPPDYVEINKEKCGE 412  
QY 241 RSQFVTVSNENKLTFRHSDQSYTDGFLAEVLSYDSDDCPQGFQRTGRCTRKELRCD 300  
DB 413 RSQFVTVSNENKLTFRHSDQSYTDGFLAEVLSYDSDDCPQGFQRTGRCTRKELRCD 472  
QY 301 GWADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDGDNSEDGSCGCPAQGTFC 360  
DB 473 GWADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDGDNSEDGSCGCPAQGTFC 532  
QY 361 SNGKCLSKSQOQCKGDKDCGSGDEASCPKXNVVTCNHTYRCLNGLCLSKGNPECDCKED 420  
XX

DB 533 SNGKCLSKSQOQCKGDKDCGSGDEASCPKXNVVTCNHTYRCLNGLCLSKGNPECDCKED 592  
QY 421 CSDGSDKDCDCGLARSFTROARVVGCTDADEGEMPMQVSLHALGQGHICGASLISPNMLY 480  
DB 593 CSDGSDKDCDCGLARSFTROARVVGCTDADEGEMPMQVSLHALGQGHICGASLISPNMLY 652  
QY 481 SAHACYIDRGRFYSPTQMTAFGLHDSQNSAPQVQERRLKRIISHPEFNDFTDYDI 540  
DB 653 SAHACYIDRGRFYSPTQMTAFGLHDSQNSAPQVQERRLKRIISHPEFNDFTDYDI 712  
QY 541 ALLELEKPAEYSWMPICLPDASHVFPAGKAIWVGWGHITQYGGTALILQGEIRVIN 600  
DB 713 ALLELEKPAEYSWMPICLPDASHVFPAGKAIWVGWGHITQYGGTALILQGEIRVIN 772  
QY 601 QTTCEMLPQOITPRMWCVGLSGVDSCQDSDGSGPLSSVEADGRIFOAGVSWDGCQAQ 660  
DB 773 QTTCEMLPQOITPRMWCVGLSGVDSCQDSDGSGPLSSVEADGRIFOAGVSWDGCQAQ 832  
QY 661 RMRPGVYTRLPFRDWIKENTGV 683  
DB 833 RMRPGVYTRLPFRDWIKENTGV 855  
RESULT 30  
AAM25628  
ID AAM25628 standard; protein; 851 AA.  
XX  
XX AAM25628;  
AC  
XX  
DT 16-OCT-2001 (first entry)  
DE  
XX  
XX Human protein sequence SEQ ID NO:1143.  
KW Human; cancer; HIV infection; human immunodeficiency virus;  
KW anti-inflammation; anti-rheumatic; anti-arthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiac; central nervous system; viricide;  
KW anti-HIV; fungicide; antitumor; cardiovascular; antineoplastic; anaemia;  
KW antileishmaniasis; haemostatic; vulvar; antileishmaniasis; eczema;  
KW dermatological; antiallergic; antistaphylococcal; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; neurotropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac amyloidosis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder.  
XX  
XX Homo sapiens.  
OS  
XX  
XX MO200153455-A2.  
PN  
XX  
XX 26-JUL-2001.  
PD  
XX  
XX 22-DEC-2000; 2000MO-US035017.  
PF  
XX  
XX 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
XX  
XX (HYSB-) HYSBQ INC.  
PA  
XX  
XX Tang YT, Liu C, Drmanac RT,  
PI WPI; 2001-457603/49.  
XX  
XX N-PSDB; AAM93569.  
DR  
XX  
XX Isolated human polynucleotides encoding polypeptides, useful for the  
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.  
XX  
XX Claim 20; Page 237; 1217pp; English.  
XX

CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 CC AAM25963. The proteins can have activities based on the tissues and cells  
 CC they are expressed in, such as: antiinflammatory; antirheumatic; anti-  
 CC antitubercular; immunosuppressive; antibacterial; endocrine; cardiant;  
 CC central nervous system; viricide; anti-HIV; fungicide; antimutagen;  
 CC cardiopulmonary; antianemic; antidiabetic; haemostatic; vulnerary;  
 CC antitubercular; osteoplastic; dermatological; antiallergic; antiaesthetic;  
 CC antidiabetic; cytoskeletal; neuroprotective; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,  
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders  
 CC  
 XX  
 SQ Sequence 851 AA;

Query Match 98.8%; Score 3745; DB 4; Length 851;  
 Best Local Similarity 99.0%; Pred. No. 4.6e-240;  
 Matches 676; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRARSLSKSFVVTSVVAFPTDSKTVQRTQDNCSPGLHARGVELMRFTTGG 60  
 DB 169 MAERVVMLPPRARSLSKSFVVTSVVAFPTDSKTVQRTQDNCSPGLHARGVELMRFTTGG 228  
 QY 61 PPDSPPYPAHACQWALRGDASVLSLTFRSFDLASCDERGSIDLTVVYNTLSPEMHALVQ 120  
 DB 229 PPDSPPYPAHACQWALRGDASVLSLTFRSFDLASCDERGHLLVYVNTLSPEMHALVQ 288  
 QY 121 LCGTYPSPSYNTLTFSSQWNLITLITNTERHPGFEATFPOLPRMSSCGRLRKAQGTFFN 180  
 DB 289 LCGTYPSPSYNTLTFSSQWNLITLITNTERHPGFEATFPOLPRMSSCGRLRKAQGTFFN 348  
 QY 181 SPYYPGHYPNIDCTWNIENVNNOHVKVFYLLBPRACGTCPKDYVEINBKYCGE 240  
 DB 349 SPYYPGHYPNIDCTWNIENVNNOHVKVFYLLBPRACGTCPKDYVEINBKYCGE 408  
 QY 241 RSQFVYVTSNSNKTIVRFHSDSYTDYGLAFLYSYSSDPGQPTCTGRCIRKELACD 300  
 DB 409 RSQFVYVTSNSNKTIVRFHSDSYTDYGLAFLYSYSSDPGQPTCTGRCIRKELACD 468  
 QY 301 GMACTDHSDELNCSGDAHQFTCKNKFCKPLFWVCDSDNDGSDSGSCPAQTFRC 360  
 DB 469 GMACTDHSDELNCSGDAHQFTCKNKFCKPLFWVCDSDNDGSDSGSCPAQTFRC 528  
 QY 361 SNGKLSKSOQCNCGDGDGSDASCPKVVVVTCTKTYRCLNGLCLSKGNPECDGED 420  
 DB 529 SNGKLSKSOQCNCGDGDGSDASCPKVVVVTCTKTYRCLNGLCLSKGNPECDGED 588  
 QY 421 CSDGSDKDCGCGLSRFTQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMWLV 480  
 DB 589 CSDGSDKDCGCGLSRFTQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMWLV 648  
 QY 481 SAACHYIDDRGFRYSDFQWTAFLGLHDQSORSAQVGERLKRILSHFPNDFTFDYDI 540  
 DB 649 SAACHYIDDRGFRYSDFQWTAFLGLHDQSORSAQVGERLKRILSHFPNDFTFDYDI 708  
 QY 541 ALLELEKPAEYSSWVRPCLPDAHVFPAGRAIWTGWHGTQGGTGAALLQKEIRVIN 600  
 DB 709 ALLELEKPAEYSSWVRPCLPDAHVFPAGRAIWTGWHGTQGGTGAALLQKEIRVIN 768  
 QY 601 QTTGENTLLPQITTPRMKCVGLSGVDSCQDSSGGLSSVADGRIFQAGVVSWDGCAQ 660  
 DB 769 QTTGENTLLPQITTPRMKCVGLSGVDSCQDSSGGLSSVADGRIFQAGVVSWDGCAQ 828  
 QY 661 RNKPGVYTRLPLFRDWIKENTGV 683

DB 829 RNKPGVYTRLPLFRDWIKENTGV 851

Search completed: November 29, 2004, 08:28:04  
 Job time : 110.142 secs

**This Page Blank (uspto)**